From the Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

APPLICATIONS OF DNA ORIGAMI ENCODED NANOSCALE PATTERNS

Ferenc Fördös

Karolinska Institutet

Stockholm 2019

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB

ISBN 978-91-7831-500-0

Karolinska Institutet

Institutionen för Medicinsk biokemi och biofysik

Applications of DNA origami encoded nanoscale patterns

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Biomedicum 1, Solnavägen 9

Onsdagen den 5 Juni, 2019, kl 09.00

av
Ferenc Fördös

Huvudhandledare:
Björn Högberg
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Biomaterials

Bihandledare:
Ana Teixeira
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Biomaterials

Fakultetsopponent:
Ebbe Sloth Andersen
Aarhus University
Department of Molecular
Biology and Genetics

Betygsnämnd:
Ulf Landegren
Uppsala University
Department of Immunology,
Genetics and Pathology

Rickard Sandberg
Karolinska Institutet
Department of Microbiology,
Tumour and Cell Biology

Per Uhlén
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Molecular Neurobiology

Stockholm 2019

Családomnak: Katinak, Ádámnak, Áronnak és Árvidnak

ABSTRACT

It was almost four decades ago when the recognition of DNA's potential use as a programmable, self-assembling building material for nanostructures led to the birth and rapid expansion of the field of DNA nanotechnology, but it was two decades later when the development of the DNA origami technique initiated the widespread use of DNA based nanoconstructs through the simplification of the design process and the reduction of the required control over the quality and stoichiometry of the assembly components by using a single-stranded "scaffold" DNA and a set of "staple" oligonucleotides that "fold" the mentioned scaffold DNA into a predesigned shape by binding different regions of the scaffold strand together. This robust approach not only permitted the construction of intricate two- and three-dimensional structures, but it also allowed the design and fabrication of molecular patterns with unprecedented accuracy as each functionalizible component's relative position in the DNA origami structure is known to nanometer precision. In this thesis we utilize the DNA origami technology's before mentioned patterning capability to create research tools for a diverse set of biomedical and biophysical applications.

In paper I. we studied the effect of different receptor ligand distributions in the ephrin/Eph signaling pathway by following the receptor activation in cancer cells stimulated with DNA origami probes displaying different, rationally designed Eph receptor ligand patterns. We found that incubation of cells with receptor ligands at shorter distance relative to each other led to significantly higher receptor activation and lower invasiveness of these cells.

In paper II. we used DNA origami to create reference samples for measuring the imaging accuracy of two of the most commonly used super resolution techniques, STED and STORM. We demonstrated that accuracy is a less biased metric for imaging faithfulness than precision and that DNA origami can be used to create a highly conserved and uniform pattern of fluorophores to measure and compare this metric for STED and STORM.

In paper III. we developed a DNA origami platform to study the photophysical behavior of two reversibly switchable fluorescent protein (rsFP) tags commonly used in microscopy in a quantitative, controlled fashion. With this system we were able to show that rsFPs at low numbers exhibit similar behavior to what was seen for them in bulk measurements, we could optimize imaging parameters more precisely and we could measure the achievable resolution using these samples. We were also able to show that some of the measured parameters scaled linearly with the amount of rsFPs making this DNA origami system a valuable calibration tool for quantitative imaging.

In paper IV. we developed a DNA origami-based optical tagging system detectable by next generation sequencing and super resolution microscopy to be used for introducing high resolution spatial information into RNA sequencing data. Using a combinatorial enzymatic approach, we were able to create a highly complex barcode library with which we successfully tagged cells and which we made compatible with one of the commonly used single cell RNA sequencing sample preparation techniques.

LIST OF SCIENTIFIC PAPERS

I. Shaw, A.; Lundin, V.; Petrova, E.; Fördős, F.; Benson, E.; Al-Amin, A.; Herland, A.; Blokzijl, A.; Högberg, B.; Teixeira, A. I. Spatial Control of Membrane Receptor Function Using Ligand Nanocalipers. *Nat. Methods* **2014**, 11 (8), 841–846.

II. Reuss, M.; Fördős, F.; Blom, H.; Öktem, O.; Högberg, B.; Brismar, H. Measuring True Localization Accuracy in Super Resolution Microscopy with DNA-Origami Nanostructures. *New J. Phys.* **2017**, 19 (2), 025013.

III. Fördős, F.; Pennacchietti, F.; Benson, E.; Högberg, B.; Testa, I. Quantitative Assessment of Reversible Photo Switchable Fluorescent Proteins for Super Resolution with DNA Origami. *Manuscript*

IV. Fördős, F.; Högberg, B. Combinatorial Enzymatic Assembly of Sequenceable, Optical DNA-Barcodes Produced by Library Cloning. *Manuscript*

CONTENTS

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
nt	Nucleotide
bp	Base pair
PCR	Polymerase chain reaction
2D	Two dimensional
3D	Three dimensional
Da	Dalton
TEM	Transmission Electron Microscopy
AFM	Atomic Force Microscopy
SR	Super resolution
TIRF	Total internal reflection
MWCO	Molecular weight cut-off
PBS	Phosphate Buffered Saline
FPLC	Fast protein liquid chromatography
ROI	Region of interest
rsFP	Reversibly switchable fluorescent protein
STED	Stimulated Emission Depletion
STORM	Stochastic Optical Reconstruction Microscopy
NGS	Next generation sequencing
NC	Nanocaliper
TBE	Tris-borate-EDTA
UMI	Unique molecular identifier
CPEC	Circular polymerase extension cloning
SIM	Structures Illumination Microscopy
SMLM	Single Molecule Localization Microscopy
RESOLFT	REversible Saturable OpticaL Fluorescence Transitions

1 INTRODUCTION

1.1 DNA as a biomolecule

DNA is one of the most fundamental biomolecules present in all known living organisms. Ever since the discovery of its abundance in living organisms[1,2] and its role as an information storage molecule[3], its chemical properties have been the focus of interest. Our knowledge about DNA's chemical nature has grown substantially since the discovery of its double helical structure. As DNA is a polymer, its chemical properties largely arise from its comprisal of repeating monomers, the nucleotides: adenine (A), thymine (T), cytosine (C) and guanine (G)[4]. These nucleotides, which consist of a five-carbon sugar molecule (2-deoxy-ribose), a purine (in the case of A and G) or pyrimidine base (in the case of T and C) and a phosphate-group, are connected through phosphodiester bonds formed by the phosphate groups (PO_4) connecting the fifth carbon (C_5) atom of a nucleotide with the third carbon atom (C_3) of the following nucleotides in the polynucleotide chain[4] (Fig. 1). Through these repeating phosphate bridges the backbone of the polynucleotide strand is formed which consists of the alternating PO_4 and sugar groups[4]. As the connection between consecutive nucleotides is asymmetrical the polynucleotide chain has an inherent directionality as well, conventionally starting at the 5' PO_4 group of the first nucleotide (5'-end) and ending at the 3' hydroxyl group of the last nucleotide (3'-end), following the direction of DNA biosynthesis[4] (Fig. 1).

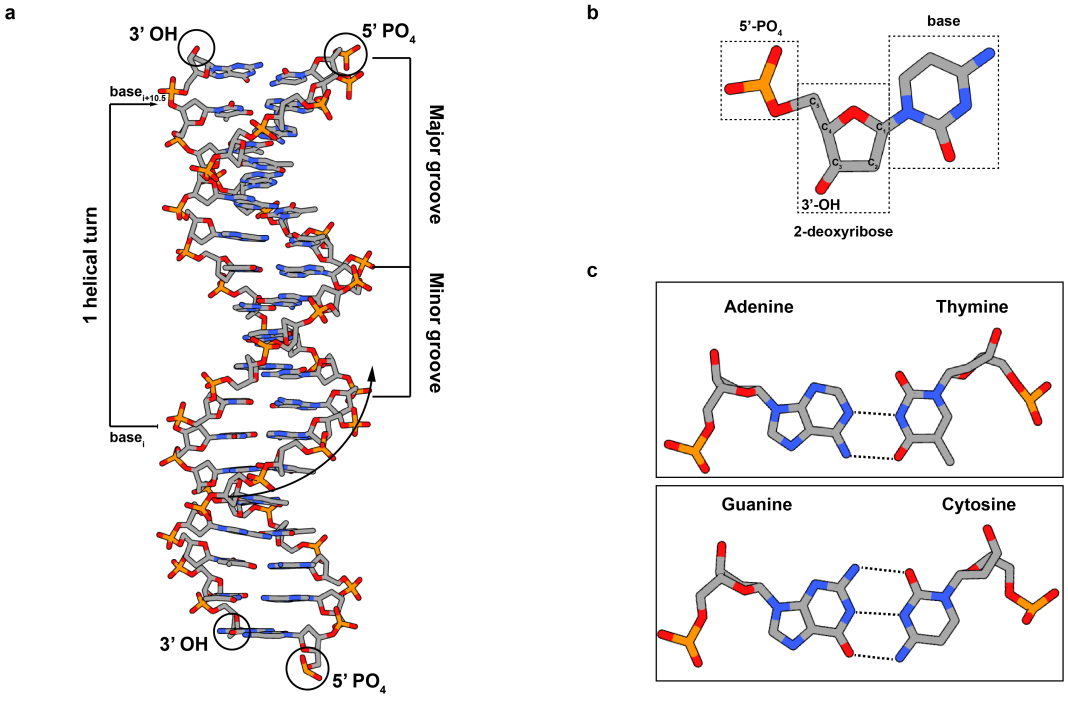

Figure 1. Structural and chemical features of B-form DNA. (a) Structure of B-DNA with its dominant structural and chemical features highlighted. **(b)** Structure of a nucleotide (Cytosine) with its main chemical groups highlighted. **(c)** Structure of nucleotides forming base pairs through Watson-Crick hydrogen bonds (dashed lines). All figures are rendered from a B-DNA crystal structure (PDB ID: 5F9I).

DNA can be found in cells mostly in a double-stranded form, where two polynucleotide strands with complementary sequences form a double helical structure in an antiparallel fashion[4], although other conformational states, such as triple-helices[5] and G-quadruplexes[6], are known to occur in vivo as well. There are a number of known distinct helical conformational variants of DNA (A-DNA, B-DNA and Z-DNA) that have been understood to have some biological relevance[7–9], however the B-form is the DNA's most common, stable conformation to be found in cells. In the B-form of the DNA the two polynucleotide strands are aligned in an antiparallel orientation, as mentioned before, with the negatively charged phosphate backbone facing outwards, towards the aqueous solvent and the nucleobases on the opposing strands facing inwards, due to their hydrophobicity[4]. The alignment of the two strands is directed by the nucleobases on the opposing strands by the formation of conserved Watson-Crick base pairs (A-T and G-C), consisting of a purine (A or G) and a pyrimidine base (T or C), through hydrogen bonds[4] (Fig. 1). Due to its distinct geometry of base pairing and sugar conformation in the backbone the B-DNA form has a set of conserved, unique structural features distinguishing it from the other conformational states of DNA. The B-form of DNA has a right-handed helical structure, with a 12 Å wide minor groove and 22 Å wide major groove found in-between the backbones[4] (Fig. 1). The B-DNA double helix has a diameter of 20 Å, a length of 3.4 Å per base pair and it takes approximately 10.5 bp for it to take one full helical turn, resulting in a 34.8˚ turn per bp[4] (Fig. 1).

1.2 DNA as a construction material

1.2.1 Synthesis of DNA oligonucleotides

The potential of DNA as a building material for biomaterial engineering was recognized almost four decades ago[10]: The energetically favorable formation of DNA's thoroughly characterized double helical structure indicates its capacity for self-assembly, while the stability and specificity of the interaction between its subcomponents, the DNA strands, can be easily tuned through the well understood chemical rules of Watson-Crick base pair formation. The major obstacle for realizing nanostructures constructed from DNA was the availability, scale (in both length and amount) and price of oligonucleotide synthesis for some time. Fortunately, because of the high demand of techniques widely used in molecular cloning, such as PCR[11] and scientific endeavors depending heavily on the use of oligonucleotides, such as the human genome project[12,13] and synthetic genomics[14], oligonucleotide synthesis became readily available and its price decreased by almost two orders of magnitude while the synthesis scale increased by more than five order of magnitude in the last three decades[15]. Presently high scales of single oligonucleotides with the maximum length of 200nt, produced by solid state chemical synthesis[16] can be purchased commercially, with other higher throughput synthetic approaches[17] appearing, which allow the parallel production of 2 million oligonucleotides improving price and production scale even further and making more demanding applications of DNA such as DNA-based data storage[18] an affordable possibility. In addition, other biological strategies have also been developed, which permit the high-scale production of

stochiometrically controlled sets of oligonucleotides with lengths not permitted by chemical synthesis by using enzymes[19] or self-cleaving DNA hairpins[20].

1.2.2 DNA nanotechnology

With the possibility of the production of oligonucleotides having a nucleotide sequence of demand the premise of DNA nanotechnology, that is, the use of DNA as a construction material for nanostructures by using a set of DNA-molecules that could self-assemble into a predesigned structure in a well-controlled way, was realized[21]. The simple underlying principle is that one can define sequence domains inside the basic structural unit of DNA, the double helix that form stable double helical structure independently from each other (Fig. 2). Following this logic, a set of oligonucleotides, designed to have a partially complementary set of sequence domains, will self-assemble into the structure dictated by their sequence complementary (Fig. 2). This assembly process is energetically favorable and only needs to be aided by an initial thermal denaturing step and the addition of positively charged cations (such as magnesium or sodium) in the case of designs with closely packed double helices to counteract the repulsion of their negatively PO_4 backbones[22].

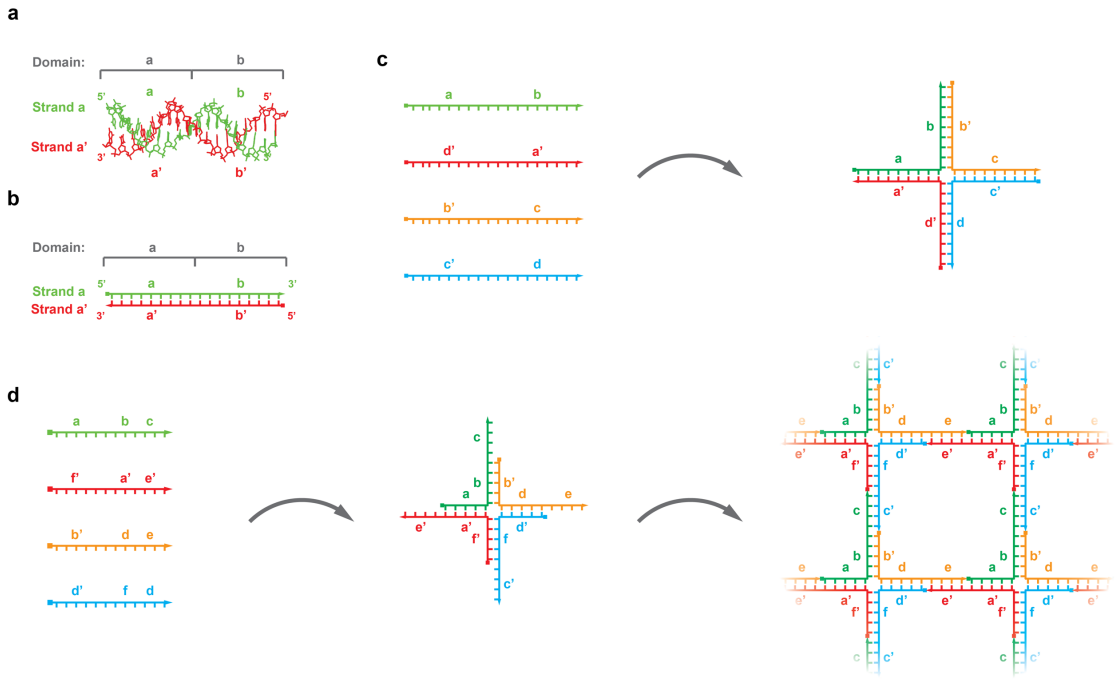

Figure 2. Demonstration of the operating principle of DNA nanotechnology. (a) Structural representation of the building block of DNA nanoconstructs, the DNA double helix, with orientation and sequence domains highlighted. **(b)** The strand diagram of the same DNA double helix with the same features highlighted. **(c)** Representation of the programmed assembly of a Holliday-junction structure driven by sequence domain complementarity of a set of two domain oligonucleotides. **(d)** Representation of the programmed assembly of a DNA-lattice structure driven by sequence domain complementarity of three domain oligonucleotides.

After the first demonstration of this principle by creating a synthetic four-armed junction[21], it was shown that by creating contacts via complementary regions ("sticky ends") between nucleic acid complexes they can be assembled into 2D crystalline lattices[23,24] and 3D objects such as cubes[25] or tetrahedrons[26] (Fig. 2). This so-called tile-based design strategy, using a set

of oligonucleotides to create nanoconstructs, has since been shown to be able to create an array of 3D objects using the same modular units[27], and its scalability in terms of complexity has been demonstrated by successfully creating a 536.4 MDa structure from 33,511 unique components[28]. Apart from fabricating specific, predesigned structures, sets of oligonucleotides with complementary sequence domains can be used to create reconfigurable dynamic systems as well for a diverse set of applications such as to define logic gates for neural network computation[29,30] or to create nanomachines, known as DNA walkers, that move in defined directions along a predefined track[31] and carry cargo[32] alternatively for molecule detection and signal amplification[33].

1.2.3 DNA origami

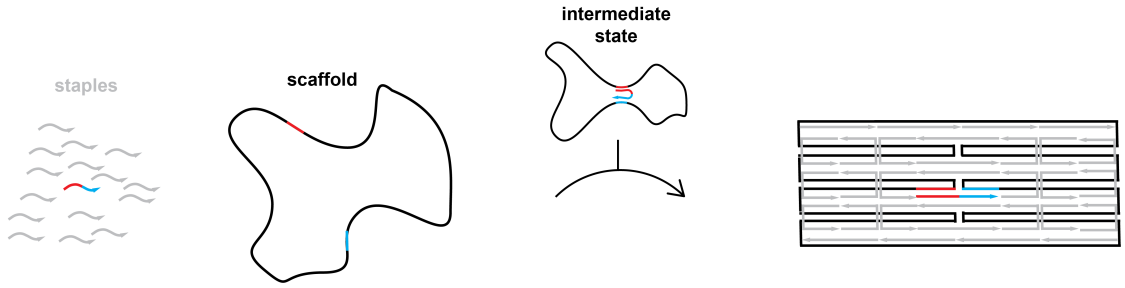

Figure 3. Producing structures with the DNA origami technique. Schematic representation of the fabrication of a DNA origami sheet structure by "folding" the single-stranded, circular scaffold DNA with a set of staple oligonucleotides containing sequence domains complementary to different regions of the scaffold molecule that they are designed to tether together (intermediate state) to create the final shape.

Another way for constructing nanostructures using DNA is to use a large circular, single-stranded DNA, called scaffold, and design a set of oligonucleotides, called staples, to "fold" the scaffold molecule into the predesigned shape by binding different regions of the scaffold with the staples containing sequence domains complementary to these regions (Fig. 3) in a so-called folding reaction using the previously mentioned heat denaturation step and cations. One of the advantages of this approach is that it is more easily scalable as it can use biologically derived scaffold DNA, using typically some version of the M13mp18 phage genome[34] as it is readily single-stranded but double-stranded sources such as phagemids[35], the lambda phage genome[36] or plasmids[37,38] can be used too. The other advantage also comes from the fact that a scaffold molecule templates the folding, which removes the need for precisely adjusting the stoichiometry of the components for folding, as only an excess of staple strands over the scaffold molecule is needed for successful folding. Additionally, the quality of oligonucleotides does not need to be as high as in the case for tile-based strategy, because a higher quality, biologically derived scaffold is used which effectively will selectively incorporate higher quality staple oligonucleotides through the greater stability of the interaction compared to low quality staples.

This strategy has been shown to enable fabrication of a great variety of complicated shapes in 2D[39] and 3D[40], with curved features[41]. It has also been demonstrated that structures can be

made dynamically reconfigurable[42–44] and used to assemble large complexes in a sequence[45] or shape-programmable way[46]. A common structural feature of classical DNA origami constructs is the parallel DNA double-helices interconnected by scaffold and staple oligonucleotide cross-overs packed in a honeycomb[40] or a less-optimal square lattice[47]. There have been a number of new strategies developed to create wireframe DNA origami structures, where double-helices are not used in the classical parallel, closely packed arrangement but rather as edges of polygon meshes used to define complex 3D[48–50] or 2D[51] structures. Structures produced with this approach show higher stability in a larger range of ionic conditions[48] and are more economical in terms of material use but have decreased rigidity[52] compared to structures produced following the classical design scheme.

1.3 Nanoscale patterns fabricated using DNA origami

1.3.1 Fabricating nanoscale patterns using DNA origami

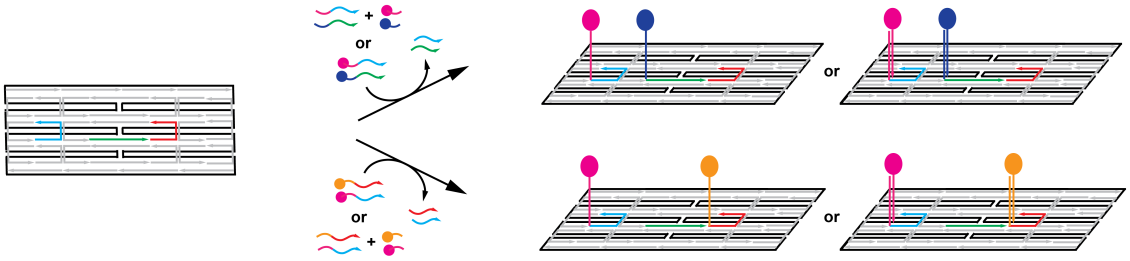

Figure 4. Creating nanoscale patterns on DNA origami structures. Demonstration of the principle for creating nanoscale patterns on DNA origami structures by replacing subsets of staple oligonucleotides with versions of them carrying different functional groups (spheres) or extra, "overhang" sequences that can be used to position functional molecules covalently attached to an oligonucleotide with complementary sequence for the overhang.

Apart from offering a remarkably efficient and relatively easy way to create complex nanoscale structures in a predesigned way using a bottom-up approach, DNA nanotechnology, and DNA origami in particular, has another highly advantageous capability: as the staple oligonucleotides are interacting with specific regions in the scaffold molecule to fold it into the predesigned shape, their relative position in the final construct is known with a nanometer precision, making DNA nanoconstructs function essentially as molecular pegboards. This can be realized by replacing staples at the desired positions with staples either carrying the functional group of desire, or in the case of heat-sensitive modifications an extra, "overhang" sequences whose complementary sequence can be covalently linked to the functional group to position it (Fig. 4.).

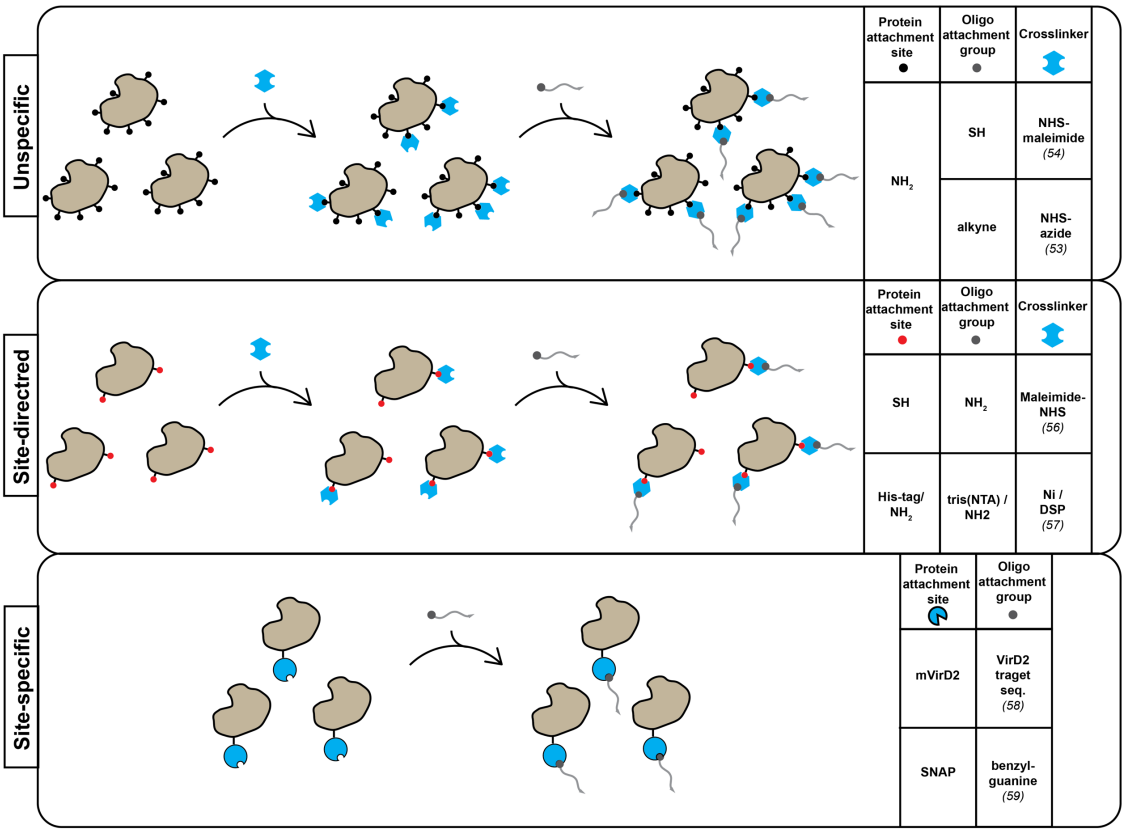

Figure 5. Strategies for creating protein-oligonucleotide conjugates. Schematic representation of the steps of three main strategies for covalently conjugating oligonucleotides to proteins with some examples for crosslinked chemical groups and crosslinking reagents used stated.

A number of strategies have been developed for the covalent modification of molecules with oligonucleotides with the main aim of achieving a high yield of modification in a positionally (attachment occurring at a designed positioned) and quantitatively (attachment of one oligonucleotide per molecule) controlled manner in reaction conditions that are not detrimental to the functional group. This is less challenging for synthetic molecules, such as fluorophores, as these can readily be synthetized with a range of modifications of desire at known positions and are generally more stable in a variety of solvents. The same is true for the oligonucleotides. However proteins, one of the popularly used biomolecules for functionalization of DNA nanostructures, are more sensitive to reaction conditions, so attachment chemistries working in near physiological conditions (aqueous buffer, pH in the stability range of the specific protein[53]) are applicable for them. The other challenge is to achieve the mentioned specificity in labelling as the chemical groups available for targeting (e.g. primary amine (NH_2) groups, sulfhydryl groups (SH)) are commonly repeated in the proteins more than one time and can play functional and/or structural roles in the proteins (Fig. 5). Still a common approach to covalently link proteins with oligonucleotides is to use bio-conjugation approaches targeting abundantly available chemical groups in the protein, such as NH_2 side groups of Lysines[54–56], and optimize the reaction conditions to achieve more controlled modifications (Fig. 5). These approaches are relatively easy to implement with high yield, with the cost of low specificity in terms of site and extent of modification. A group of targeted attachment strategies achieve a more controlled, site-directed modification of proteins by targeting amino acids of low

abundance (e.g. Cysteines)[57] or by using Histidine-tags[58] to direct the attachment of oligonucleotides to protein via metal-coordination, but they still do not have absolute site specificity and they are more complicated to implement. Finally, there are techniques that achieve total site-specificity utilizing DNA-[59] or small-molecule binding (SNAP)[60] tags fused to the target proteins, however these techniques cannot be utilized on native proteins and they use relatively large tags (≥~20 kDa) for the attachment, which can be limiting for some applications (Fig. 5). Nevertheless, with these techniques and other approaches being currently developed[61], DNA origami is becoming widely used platform for creating molecular patterns with uniquely high, nanometer precision.

1.3.2 Biological applications of DNA nanostructures

Because the size of most of the macromolecules inside cells are within the nanometer range the ability to create patterns of biomolecules with nanometer precision with DNA origami is unquestionably valuable to answer a plethora of scientific questions, which were challenging to do before. One of the subfields DNA origami has made contributions to is biotechnology, where it has been shown to be a useful platform to increase the efficiency of multi-enzyme chemical reactions by positioning members of enzymatic cascades in arrays[56], or to modulate enzymatic activity by controlling accessibility of enzymes[54]. DNA origami has also contributed to our understanding of the behavior of motor enzyme ensembles through using DNA origami-based synthetic cargos functionalized with opposite-polarity motor enzymes[60]. The other biological field where molecular patterns are of known importance is immunology, as recognition of molecular motifs form an important part of both innate and adaptive immunity[62]. In the latter case the molecular patterns can be important for antibody-mediated immune response against extracellular pathogens in particular, for this reason a DNA origami system has been used to study the effect of the dimensions and configuration of antigen-patterns on the binding efficiency of antibodies[63]. Outside of the field of molecular biology, DNA origami structures are a focus of interest in the field of drug delivery as well. Apart from being able to functionalize DNA structures with therapeutic agents, such as DNA intercalating chemotherapeutic agents[64,65], as it have been mentioned before they can be designed to have actuatable, dynamic features. It has been demonstrated that by coupling these switching mechanisms to target (e.g. tumor antigen) specific markers one can make drug filled molecular robots which only release their cargo in the proximity of target cells[55,66]. One limitation hindering the broader application of DNA origami drugs in therapeutics is the instability of DNA constructs in biological fluids, because of the low concentration of cations or the presence of nucleases. Both of these problems are being addressed by the community, the former by stabilization of the structure by covalent crosslinking[67] or replacement of cations with oligolysine[68], the latter the protection of the structures with PEG-[68] or lipid-encapsulation[69], making the future clinical use of DNA origami based drugs a possible reality.

1.3.3 Imaging applications of DNA nanostructures

DNA origami's capability for creating programmable nanoscale patterns of a plethora of different molecules made it a popular technique for imaging applications as well. One of the

subfield DNA origami has gained a widespread use in is super resolution (SR) microscopy. Classically light microscopy's resolution was physically limited to roughly half of the illuminating lights wavelength (corresponding to a minimum of ~250 nanometers) by the diffraction of light, meaning that the molecules in the sample being imaged that are closer to each other than this distance cannot be detected as two objects[70]. This posed a major obstacle in understanding the organization and interaction of biomolecules in cells as the typical sizes of the imaged molecules are well below this size range[71]. A number of techniques have been developed in order to solve this problem by achieving sub-diffraction resolution with light microscopy, which can be grouped into two main categories based on their approach to achieve this. The targeted switching methods (e.g. STED[72], SIM[73,74], RESOLFT[75]) do this by spatially separating the emission of the fluorescently labelled molecules in close proximity to each other in the sample by selectively activating (SIM[73,74]) or keeping active (STED[72], RESOLFT[75]) a subset of fluorophores residing in sub-diffraction limited volume (Fig. 6). The techniques belonging to the second group, called Single Molecule Localization Microscopy (SMLM) methods (e.g. STORM[76,77], PALM[78,79]), separate the emission of fluorescent markers in the samples temporally by making only a random, sparse subset of fluorescent molecules in the sample emit photons ("blink") at any given time and determining their position with sub-diffraction precision (Fig. 6). DNA nanotechnology also made an impact on the field of SMLM techniques by creating an imaging technique called DNA-PAINT[80] that uses fluorescently labeled oligonucleotides as probes to transiently bind to the imaged target molecules labeled with complementary oligonucleotides to create the SR image. The fact that the generation of the signal is based on the predesigned, tunable interaction of oligonucleotides makes it possible to create false colors by altering either the used imaging oligonucleotides' sequences[81] or by tuning the strength of their interaction[82], additionally it permits a precise way for quantitative imaging[83].

With the breaking of the diffraction limit new reference samples were needed for testing and further improving the experimental capabilities of these methods. DNA origami quickly became a popular platform to create such samples, displaying emitters at a predesigned, sub-diffraction distance in a highly uniform and repeatable manner for 2D[84] and 3D[85] SR imaging methods (Fig. 6) using fluorophores[84,85] or fluorescent proteins[86]. Apart from creating calibration standards for SR imaging techniques DNA origami has also been demonstrated to be a good platform for creating highly multiplex labeling probes. By positioning a set of different fluorophores on DNA origami constructs geometrical information has been used to produce a probe-library[87] with orders of magnitude higher complexity compared to what is achievable with conventional, spectrally encoded tagging systems (Fig. 7). The probes have been shown to be applicable for SR imaging as well as standard epifluorescence microscopy. The control over the functional groups displayed by DNA origami structures also permits to discretely tune the functionalization extent of these nanoconstructs. This can also be exploited to create sub-diffraction sized probes with predesigned quantities of fluorescent molecules

distinguishable by their fluorescent intensity, that can be used for intensity calibration or as another system for molecular barcoding[88].

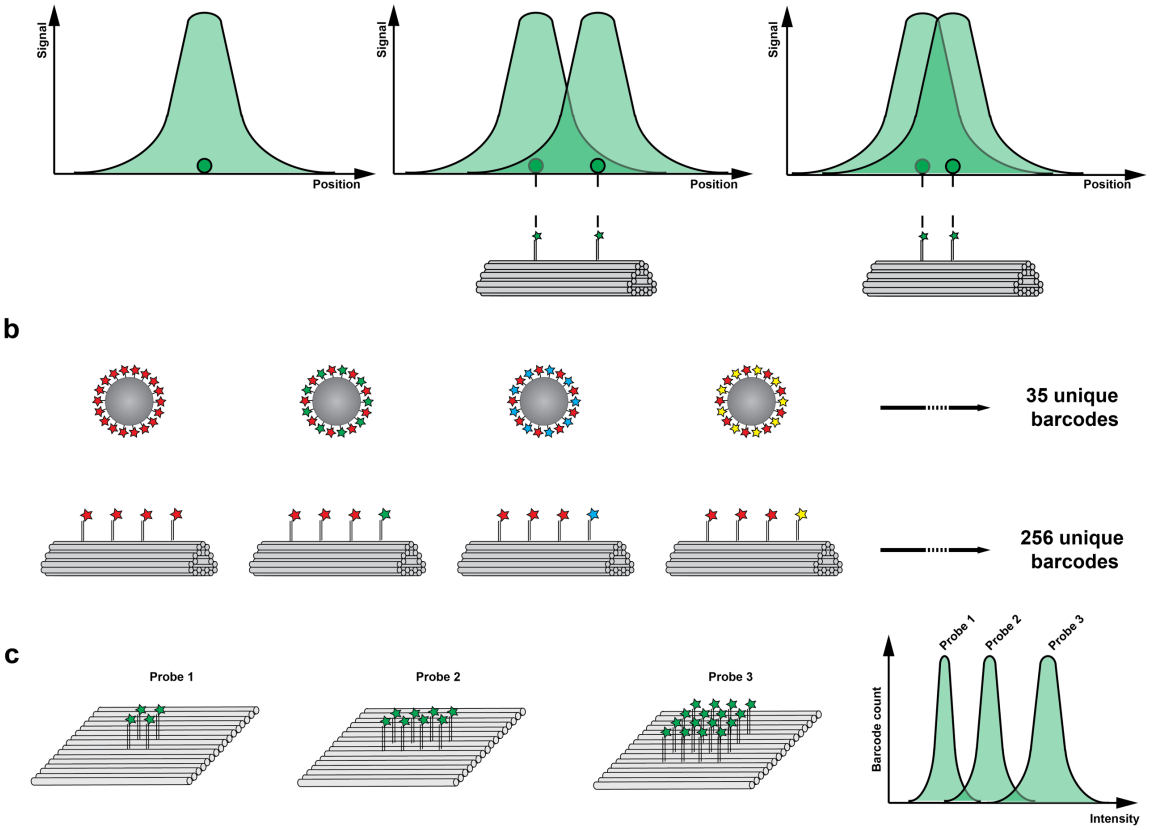

Figure 6. Imaging application of DNA origami structures. (a) Using DNA origami structures for testing the resolution of light microscopy techniques by placing fluorophores at well-defined distances **(b)** Fabrication of geometrical optical barcodes with DNA origami (bottom row) which compared to other classical approaches (top row) permits the production of orders of magnitude larger label libraries using the same number of distinct probes. **(c)** Creating fluorescent probes (left) distinguishable by their intensity (right) using DNA origami technique's capability for discretely tuning the functionalization extent of the designed structures.

2 AIMS

The aim of the work presented in this thesis is to create research tools for biomedical and biophysical applications using the DNA origami technique's previously described capability for creating nanoscale molecular patterns to address research questions that were previously challenging to answer. The specific aims of the papers presented in this thesis are:

Paper I. – To investigate the effect of nanoscale ligand distribution on receptor activation in the Eph-receptor/Eph signaling system by stimulating EphA2-receptor expressing cancer cells with rationally designed EphA2-receptor ligand patterns displayed on DNA origami structures.

Paper II. – To probe and compare the capabilities of the two most commonly used SR imaging techniques, STED and STORM, in terms of their true localization accuracy using DNA origami structures functionalized with fluorophores at predesigned positions with a predetermined nanometer scale distance between them.

Paper III. – To study the single molecule photophysical behavior of reversibly photo switchable fluorescent proteins (rsFPs) used in RESOLFT SR imaging and to test how these properties scale with probe quantity and how that influences the achievable resolution by controlling the number and the position of preprogrammed amounts of rsFPs on DNA origami probes.

Paper IV. – To create a sequence encoded, optical tagging system for cells using DNA origami with the optical code detectable by microscopy and next generation sequencing to develop a spatial transcriptomics technique building on current single cell RNA sequencing methods.

3 MATERIALS AND METHODS

3.1 Design of DNA origami structures

For most of the presented works the main requirement for the structures was the fidelity of the produced structures to the designs to be able to create molecular patterns with high precision and as most of the patterns used were in 2D, rod-like structures were used for the majority of the cases. For the design we followed the classical DNA origami strategy relying on parallel double-stranded helices packed in a honeycomb lattice (18 helices in the work presented in paper I.-III. and 12 helices in paper IV.) to achieve high rigidity with a minimum use of material. We used the caDNAno software[89] for this (Fig. 7).

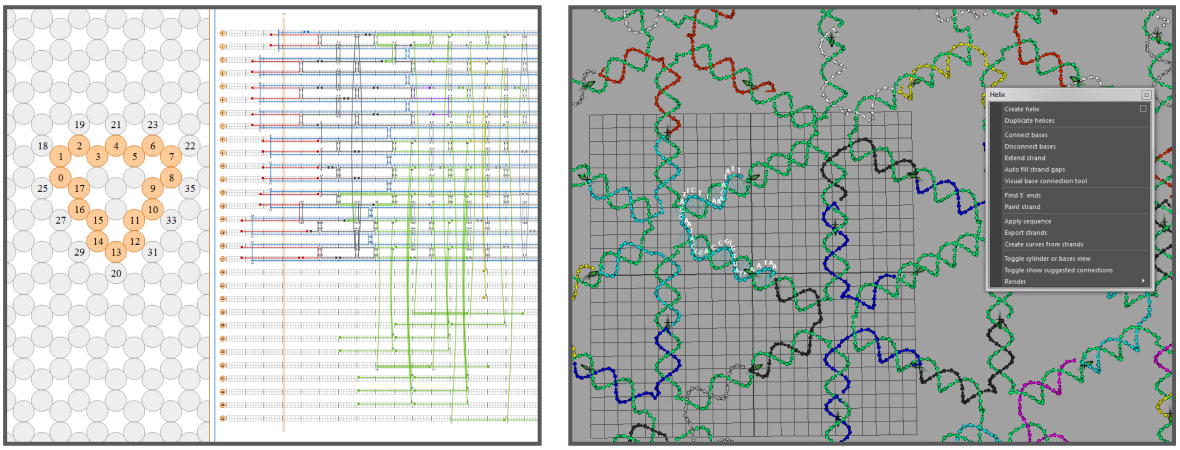

Figure 7. Softwares used for designing DNA origami structures. Crop out of the user interface of the two softwares used to design the DNA nanostructures: the caDNAno software (left) used for designing classical DNA origami structures showing the design of the 18 helix bundle structure, and the vHelix software (right) used for designing the wireframe structure displaying the user interface and a close up of the polyhedral sheet structure.

The general procedure that we followed was that we first decided the dimension of the designed patterns (number of molecules per position and distances between positions) and then calculated the number and length of helices to accommodate this pattern based on the known geometrical rules of DNA origami structures based on a honeycomb lattice (0.34nm per base, 21 nt for every two helical turn). After we created the parallel helices with the calculated length in the caDNAno software we created connection points between them, or scaffold crossovers, in order to define the scaffold's path inside the structure. We then used the caDNAno software's *auto staple* feature to create staples strands connecting the designed scaffold path. We then manually introduced the staple break-points to create staple strands with the length between 21 nt and 60 nt. At the terminal part of the structures either unpaired, single-stranded scaffold regions or staples with single-stranded protrusions were designed for counter acting dimerization of structures driven by base-stacking. Functional sites (containing anchoring sites for patterned molecules, fluorophores or biotin groups for surface immobilization) were designed by creating staple break-points at positions where the current base of the staples was facing "outside" from the structure. Based on the length of the helices in the structure a scaffold with an appropriate length (p7560 for the rod structures use in paper I.-III.) was chosen and its sequence was applied to a randomly chosen starting point in the scaffold path of the structure

to calculate the sequences of the designed staple oligonucleotides using the caDNAno software's *add seq* feature. The staple sequences where then processed in Microsoft Excel to add functional sites to the selected staples and the staple oligonucleotides were ordered from Integrated DNA Technologies. This approach was slightly altered in the work presented in paper IV. The barcode structure's design was preceded by the design of the insert, encoding the optical code of the barcode structure, and the custom scaffold containing it. The length of the helices in this case was calculated to accommodate the full length of this insert and the number of helices was reduced to 12 to decrease the structure's material use without the compromise of rigidity. The other alteration in the design flow concerned the routing of the scaffold through the structure, as the scaffold insert was designed to run through the top helix of the structure uninterrupted by internal scaffold crossovers and containing unpaired, single-stranded regions at the positions (the color encoding sites and the unique molecular identifier (UMI) site) with variable sequences.

For paper III. we also utilized a newer strategy[51] to create a polygonal wireframe sheet structure (PGS) allowing us to create a high number of functionalization sites (192 possible functionalization sites) with higher neighbor-to-neighbor distances (~6 nm distance between adjacent sites) that is otherwise not possible with the classical approach in order to be able to achieve a high functionalization yield that is less limited by the steric hindrance between the introduced molecules. For this design a triangulated wireframe mesh[51] representing the target design was created in Autodesk Maya and exported in the STL format and the exported file was converted to the ASCII PLY format using the software MeshLab. The BSCOR software package (available at www.vhelix.net) was used to route a scaffold path through the mesh and convert it to a DNA origami geometry. This was imported to the DNA nanostructure design software vHelix[48] (Fig. 7). In vHelix, the feature "auto-fill strand gaps" was used to pad gaps at strand junctions with unpaired nucleotides to reduce strain. Every helix in the DNA structure featured a staple breakpoint that was initially positioned in the center, and oriented randomly. These staple breakpoints where moved along the helix in order to point "upwards", for protein attachment sites, and "downwards", for surface immobilization sites, away from the plain of the structure. The sequence of the scaffold p8634 was applied to the structure, generating the sequences of the complementary staple strands. The staple sequences where then processed in Microsoft Excel to add functional sites to the selected staples and the staple oligonucleotides were ordered from Integrated DNA Technologies.

3.2 Production of scaffold ssDNA

3.2.1 Production of p7560 and p8634 scaffolds

For producing the single-stranded scaffold molecules presented in paper I.-III. (p7560 and p8634) the standard protocol established in the DNA origami field was used. Large scale bacterial culture in 2xYT medium (containing 5mM $MgCl_2$) was produced from a clonal origin using an E. coli strain (JM 109 and K91) carrying the F gene (F^+, Hfr and F' strains) required for infection by the M13mp18 phage. This culture was then infected at the exponential stage of its growth with a version of M13 phage containing an insert in its 7249 nt long genome to

increase the size of the produced single-stranded DNA (ssDNA) to the desired size (7560 nt or 8634 nt). After the amplification of the phage during a 4 to 5 hours long incubation at 37°C the bacteria was removed from the culture with centrifugation. The phage particles were subsequently isolated by PEG precipitation and the ssDNA was extracted by the removal of the coating of the phage by alkaline lysis followed by a purification step utilizing ethanol precipitation. The resulting purified scaffold ssDNA pellet was resuspended and stored in 10 mM TRIS buffer (pH 8.5).

3.2.2 Combinatorial production of the scaffold library

3.2.2.1 Combinatorial production of scaffold insert library

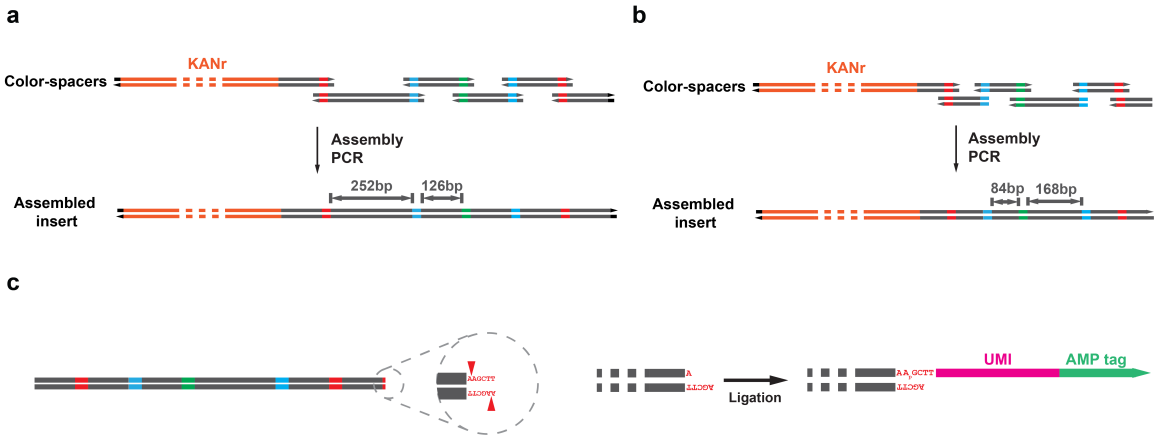

Figure 8. Combinatorial production of the scaffold insert library. (a-b) The barcode insert libraries were constructed from color-spacer sequences, consisting of constant spacer sequences flanked with color sites containing one of the four possible DNA-PAINT docking site pairs and an overlapping region to the adjacent spacer, using combinatorial assembly PCR . The insert libraries contained a KAN_R gene (orange) for positive selection of the transformed bacteria. The inserts were designed to be asymmetric to enable the distinction of order in the imaging by using one large spacer (**(a)** 252 bp for barcode insert library v1 and **(b)** 168 bp for the barcode library version 2) together with five shorter spacer segments (**(a)** 126 bp for the barcode library v1 and (**b)** 84 bp for barcode library version 2). **(c)** Labeling scheme of the barcode insert (v2) molecule with a UMI sequence through a ligation reaction for counter-acting UMI tagging promiscuity.

A new approach was developed to produce the custom scaffold used for the barcode structures presented in paper IV. The two versions of barcodes we produced contained an insert containing the sequences ("color sites") encoding the optical code of each barcode (Fig. 8). The insert molecule was designed to contain five color sites, each consisting of a distinct pair of docking sites for DNA-PAINT imaging, arranged asymmetrically to permit order identification, resulting in a total of 1024 possible color permutations (Fig. 8). Conserved spacer sequences and a set of assembly bridge primers, containing a pair of DNA-PAINT docking site sequences[90] used and a 19nt overlapping sequence to the adjacent spacers, were used for assembling the inserts using a combinatorial PCR approach (Fig. 8). The insert molecules were fused for both barcode versions to a Kanamycin resistance gene in order to enable positive selection of the insert carrying constructs during cloning (Fig. 8). The assembled inserts were cleaned using agarose gel extraction and cloned into M13mp18 vector. In the case of barcode library version 1 five 126 bp and one 252 bp long spacer were produced using PCR from pUC18 vector. Three DNA-PAINT docking site sequences (P1, P2 and P4) were used to create the four color sites (Blue (B): P1-P1, Green (G): P2-P2, Red (R): P4-P4,

Yellow (Y): P2-P4). The assembled insert was cloned into M13mp18 using ligation of linear M13mp18 and the insert molecule, containing complementary restriction enzyme digested sticky ends. For the barcode library version 2 we implemented some changes to the insert library production. We used shorter, synthetic spacers (five 84bp and a 168bp long) and changed the position of the long spacer to allow the use of the higher throughput Illumina Miseq sequencing platform for the NGS characterization of the library. Additionally, we used four distinct DNA-PAINT docking sites in the color-spacer sequences (Cyan (C): P1-P2, Green (G): P2-P2, Purple (P): P4-P4, Yellow (Y): P6-P6) and implemented a strategy in order to compress the information of a barcode sequence into a tag readable by short read length platforms more commonly used for RNAseq by using a 25nt long unique molecular identifier (UMI)[91] that we linked to the assembled inserts in a ligation reaction using T4 ligase (Fig. 9). We also changed the cloning strategy to a PCR based, circular polymerase extension cloning (CPEC)[92] method in order to avoid unintended cuts made by restriction enzymes in the UMI sequences randomly containing recognition sites.

3.2.2.2 Production of scaffold library

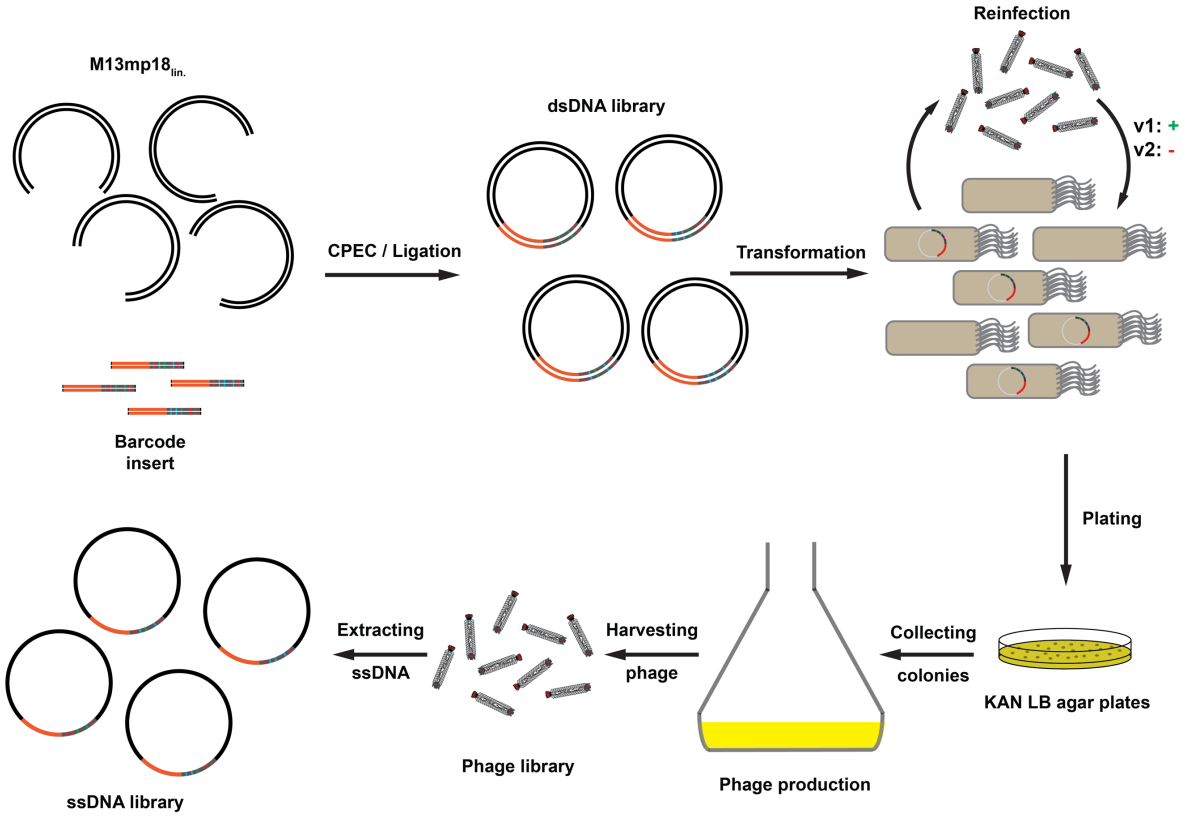

Figure 9. Production of ssDNA scaffold library. Schematic representation of the steps for the scaffold library production process: Barcode insert library was cloned into the M13mp18 vector and was transformed into bacteria (F^+ for library version 1 and F^- for library version 2). The transformed bacteria was then selected for the construct by growth on Kanamycin containing plates. Colonies were collected and transferred into high volume cultures for large-scale production of the phage library. Established ssDNA extraction protocol was used for the isolation of the ssDNA library from the culture.

For barcode library version 1 (v1) the ligation reaction was then used for chemically transforming a phage competent (F^+) K91 bacteria and the transformation mixture was plated out before phage production on Kanamycin containing LB-agar plates (Fig. 9). This allowed us to control and estimate the library diversity through the number of colonies and also

counteracted to some extent the domination of the library by random clones. We then collected and transferred the colonies into overnight cultures in 2xYT (50 µg/mL KAN, 5 mM $MgCl_2$) for the production of the phage library. The ssDNA scaffold library was extracted from the culture following the procedure described earlier.

For the second version (v2) of the barcode library we altered the protocol to further increase the yield and diversity of the library. In this case the CPEC mixture was used for transforming a phage non-competent (F^-) ultra-, electrocompetent bacterial strain (MegaX DH10B $T1^R$) and the transformation mixture was similarly plated out before phage production on Kanamycin containing LB-agar plates. The use of a phage non-competent strain was adapted in order to prevent the skewing of the library distribution by exponential amplification of random constructs through reinfection during the cloning (Fig. 9). We collected the colonies from the plates and produced the phages following the same procedure that we used for barcode library version 1.

3.3 Folding, purification and characterization of DNA origami structures

Structures used in the presented work were produced using a folding reaction. The folding reaction was carried out by mixing the respective scaffold molecule (18HB: p7560, PGS: p8634, Barcode structure v1/v2: scaffold library v1/v2) at 20 nM concentration with their respective synthetic staple oligonucleotide mixture (containing all staple oligonucleotides needed for structural integrity and the functionalization of the structures) at 100 nM individual concentration in a Mg^{2+} folding buffer (5 mM TRIS, 1 mM EDTA). The optimal Mg^{2+} concentration was determined for all structures individually (18HB: 13 mM, Barcode: 10 mM) in an initial Mg^{2+} folding screen. The polygonal structures were folded in 1X PBS as they do not contain closely packed helices needed to be shielded by bivalent cations, additionally we observed higher-quality folding with less dimerization compared to folding in Mg^{2+} folding buffer. The folding mixture was then subjected to an overnight temperature ramp. We used a shorter ~16 hours program for the 18HB structure consisting of an initial denaturation step at 80 °C, followed by a slow cooling step from 80 °C to 60 °C over 20 minutes and a final 15.5 hour long cooling step from 60 °C to 24 °C. For the wireframe sheet structure and for the barcode structure we used a longer ~20 hour program starting with a 65 °C denaturation step for 15 minutes followed by a quick transition to 60 °C and finishing with a slow cooling from 60 °C to 40 °C over 20 hours and ending in a quick transition to room temperature. The excess of the staple oligonucleotides was removed by washing the folding reaction in 100 kDa MWCO 0.5 ml Amicon centrifugal filter columns using Mg^{2+} storage buffer (10 mM Mg^{2+}, 5 mM TRIS, 1mM EDTA). The concentration of purified structures was determined by measuring the absorbance of the purified folding mixture at 260 nm with the Nanodrop instrument. The quality of folding was assessed by running the structures in 2% agarose gels (0.5X TBE, 10 mM Mg^{2+}, 0.5 mg/mL Ethidium bromide (EtBr)) in an ice water bath for 2 at 90 V or 4 hours at 70 V respectively. Classical origami rod structures are generally compacted upon folding, resulting in an increased migration speed compared to the scaffold molecule (Fig. 10). In contrast to this, wireframe structures, and the polygonal sheet structures in particular, show an inverse behavior: they tend to migrate slower than their respective scaffold molecule due to

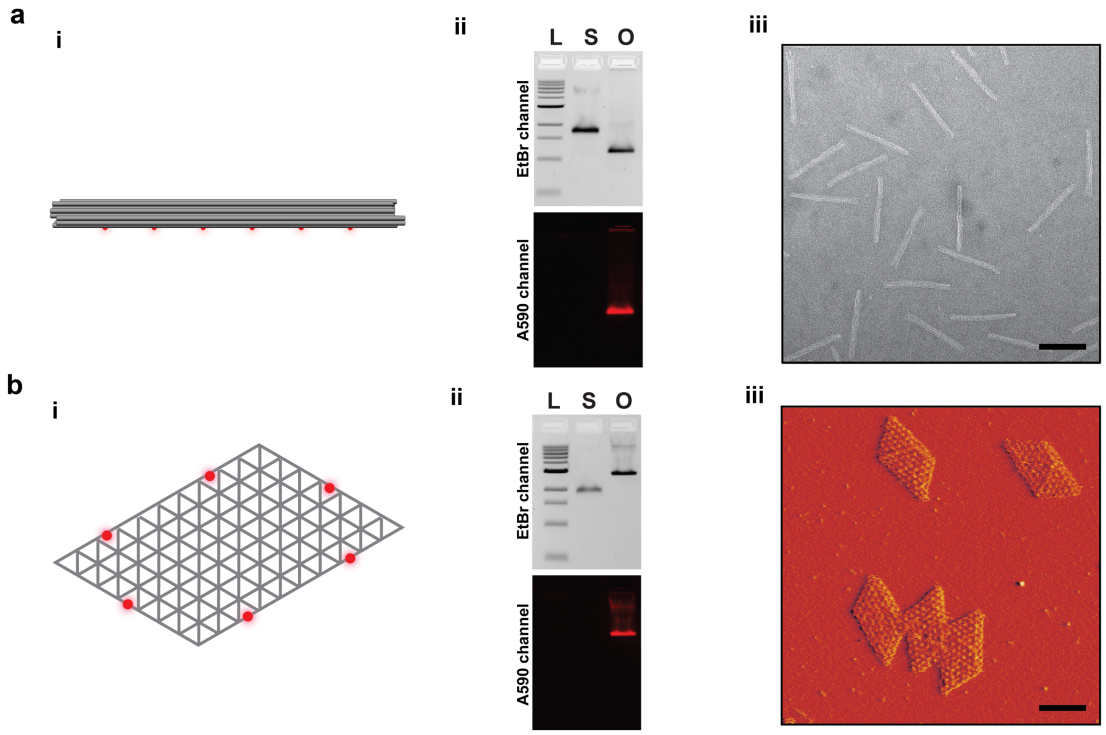

Figure 10. Production and characterization of DNA origami structures. (a/i) Schematic representation of a rod like structure, exemplified by the 18HB structure used in paper III. with DNA double helices represented by cylinders and the displayed Atto 590 fluorescent markers as red spheres. **(a/ii)** Quality control of folded and purified 18HB structures using 2% agarose gel electrophoresis showing an increased migration speed of the DNA origami band (O) compared to the scaffold band due to the compactness of the structures and the successful incorporation of the Atto 590 tags. **(a/iii)** Negative-stained TEM images of the folded and purified 18HB structures (scale bar = 100 nm). **(b/i)** Schematic representation of the polygonal wireframe sheet structure (PGS) used in paper III. with DNA double helices represented by edges and the displayed Atto 590 fluorescent markers as red spheres. **(b/ii)** Quality control of folded and purified PGS structures using 2% agarose gel electrophoresis showing a decreased migration speed of the DNA origami band (O) compared to the scaffold band due to the increased surface of the structures and the successful incorporation of the Atto 590 tags. **(b/iii)** AFM images of the folded and purified PGS structures (scale bar = 100nm)

their increased surface area (Fig. 10). This electrophoretic assay is also useful for checking the incorporation of designed fluorescent markers. More detailed information can be gathered about the folded constructs structural integrity by using high resolution imaging techniques. For imaging 3D structures negatively-stained transmission electron microscopy (TEM) was used. Samples were prepared by incubating a drop of solution with folded structures on top of glow-discharged Carbon-coated Formvar grids for 20 seconds and staining the grid with 2% aqueous uranyl formate solution after sample drop was blotted off. A FEI Morgagni 268(D) TEM was used at 80 kV to collect images with nominal magnifications between 18000-44000X. The imaging of the 2D polyhedral wireframe sheet was performed using atomic force microscopy (AFM). DNA origami sample diluted to 0.5 nM in Mg^{2+} storage buffer (10 mM $MgCl_2$, 5 mM TRIS, 1 mM EDTA) was incubated on top of freshly cleaved mica surface inside a fluid-imaging cell, fabricated by gluing a mica disc to a microscope slider and gluing a plastic ring around the mica disc. After 30 seconds, $NiSO_4$ with an end concentration of 1.4 mM was added to the surface and incubated for 4.5 minutes. After this the sample was washed by with 1 ml of imaging buffer (10 mM NaCl and 1 mM $NiSO_4$). 1.5 ml of imagining buffer was added to the fluid cell, and the sample placed in a JPK nanowizard ultra 3 atomic force microscope. The structures were imaged using a Bruker Scanasyst fluid + cantilever in AC mode.

3.4 Production of protein-oligonucleotide conjugates

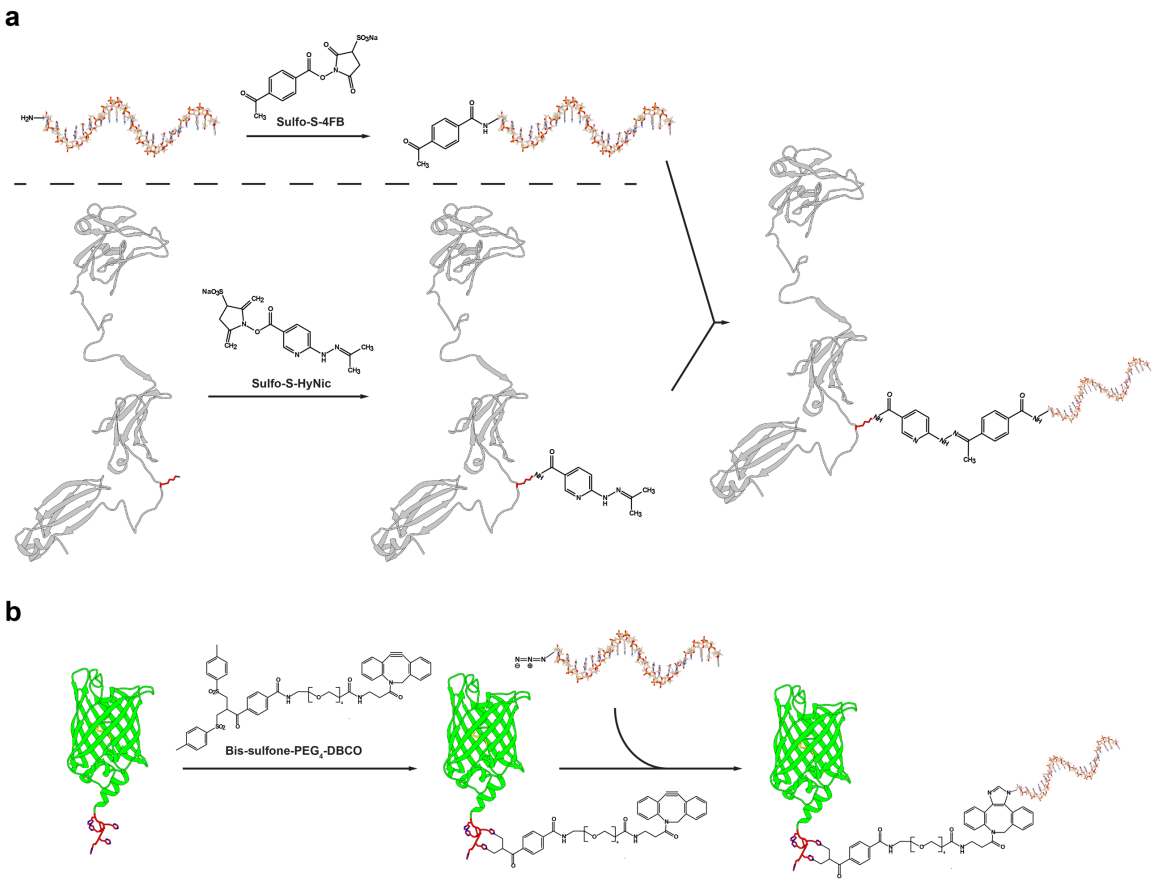

Figure 11. Workflow of chemical strategies used to create protein-oligonucleotide conjugates. (a) Conjugation of Ephrin-A5-Fc chimera with 3' NH_2-modified oligonucleotide using a hydrazide/hydrazone click chemistry: First the modified oligonucleotide was reacted with Sulfo-S-4FB reagent targeting the NH_2-group and the Ephrin-A5-Fc chimera with Sulfo-S-HyNic reagent targeting the primary amine groups in Lysine side chains. Modified Ephrin-A5-Fc and anchoring oligonucleotides were then mixed to create the conjugates through the formation of a hydrazone bond between the introduced crosslinking groups. **(b)** Conjugation of rsFP with 3' azide-modified oligonucleotide using a site-specific alkyne/azide click chemistry: First the rsFP were reacted with Bis-sulfone-PEG_4-DBCO reagent that selectively reacts with imidazole groups in the His-tag present in the protein. Modified rsFP and anchoring oligonucleotides were then mixed to create the conjugates through a Cu-free click reaction between the azide group in the oligonucleotide and the DBCO group in the rsFP bound crosslinker.

3.4.1 Conjugation of ephrin-A5-Fc chimeras using hydrazide/hydrazone click chemistry

Eph receptor ligand used in paper I. was conjugated to the anchoring oligonucleotide used for patterning on the DNA origami structure using a three-step hydrazide-hyrazone click chemistry (Fig. 11). First, the 21 nt long, 3' NH_2-modified anchoring oligo was reacted with ~45-times molar excess of Sulfo-succinimidyl-4-formylbenzamide (Sulfo-4-S-FB) in reaction buffer (0.5 mM EDTA buffer, pH 8.0) at room temperature for 1 hour before the addition of an equal amount of Sulfo-4-S-FB reagent and incubation for 1 more hour. The functionalized oligonucleotide was purified and buffer exchanged into conjugation buffer (1X PBS, pH 6.0) with 5 kDa MWCO Vivaspin spin filter columns and stored at 4 °C until further use. Recombinant human ephrin-A5-Fc chimera were used as Eph-receptor ligands that were reacted with Sulfo-succinimidyl 6-hydrazinonicotinate acetone hydrazine (Sulfo-S-HyNic), that selectively reacts with primary amines in the protein, mostly found in the Lysine side

chains. The monomer of the protein contained 22 accessible Lysines out of which only 5 were found in the ephrin domain, the 17 Lysines resided in the functionally less important Fc part of the chimeric protein, making it more probable target for the modification. Eph-A5-Fc was reacted in reaction buffer (1X PBS, pH 7.4) with Sulfo-S-HyNic for 2 hours at room temperature with ~10 time molar excess of reagent over protein. The modified proteins were cleaned and buffer exchanged into conjugation buffer (1X PBS, pH 6.0) with 7 kDa MWCO Zeba Spin desalting columns. Finally the 4FB-modified anchoring olignucleotides were mixed and reacted with the HyNic-modified EphA5-Fc for 2 hours at room temperature with ~10 molar excess of oligonucleotides over the protein. The excess of oligonucleotides was removed and the conjugates were buffer-exchanged into storage buffer (1X PBS, pH 7.4) using 50 kDa MWCO 0.5 ml Amicon centrifugal filter columns. The conjugation efficiency was evaluated by measuring the concentration of the protein with the Bradford assay and the concentration of the formed hydrazone bond with absorbance at 350 nm for the conjugates. With the optimization of reaction conditions, a conjugation yield of 0.9-1.3 oligonucleotide per ligand was achieved.

3.4.2 Conjugation of rsFPs using a site-specific alkyne/azide click chemistry

The transgenically produced rsFPs (rsEGFP2 and rsEGFP(N205S)) carrying a His6-tag used in paper III. were conjugated to azide (N_3)-modified anchoring oligonucleotide by using a site-specific alkyne/azide click chemistry using Bis-sulfone-PEG_4-Dibenzocyclooctyne (BS-PEG_4-DBCO), which selectively reacts with imidazole rings of the Histidine tags[93](Fig. 11). rsFP were reacted in reaction buffer (1X PBS, pH 6.3) with a 10-time molar excess of BS-PEG_4-DBCO for 4 hours at 37°C. DBCO-modified rsFP was purified and buffer exchanged into conjugation buffer (1X PBS, pH 7.2) with 7 kDa MWCO Zeba Spin desalting columns. Finally, the DBCO-modified rsFP was reacted with the 21 nt long, N_3-modified anchoring oligonucleotide in an overnight reaction at room temperature with the oligonucleotide in 10-times molar excess over the protein. The excess of oligonucleotides was removed by size-exclusion chromatography using a Superdex 75 10/300 GL column (GE Healthcare) coupled to an ÄKTA FPLC instrument (GE Healthceare) with the 280 nm UV measurement decoupled. The fraction containing monomeric rsFP-oligonucleotide conjugates were concentrated using 50 kDa MWCO 0.5 ml Amicon centrifugal filter columns. Concentrations of conjugates were determined by measuring the absorbance at the absorbance maximum of the rsFPs.

3.5 Production and characterization of functionalized DNA origami structures

3.5.1.1 Production of functionalized DNA origami structures

Protein-oligonucleotide conjugates were produced as describes earlier. Oligonucleotides linked to other functional groups (fluorophores: Atto 590, Alexa Fluor 488, CAGE 552; attachment groups: N3, biotin) were purchased commercially. Thermostable groups that had non-variable positions (e.g. biotin-groups in surface anchoring oligos) in the structures were designed as modified staple-oligonucleotides with a double-thymidine spacer between the staple terminus and the functional group and were included in the folding reaction together with other staple

oligonucleotides. Thermostable modifications with alterable positions (Alexa Fluor 488, CAGE 552) were also included in the folding mixture along with the staple-nucleotides with complementary overhanging sequences used for patterning them. Thermosensitive modifications (ephrin-A5-Fc- and rsFP-oligonucleotide conjugates) were added to structures after folding and purification. First, structures with protruding anchoring staple oligonucleotides at designed positions were folded and purified as described before. Complementary oligonucleotides conjugated to the functional groups were added to the structures with a 2 to 4 times molar excess of conjugates over anchoring sites. The mixture was then subjected to a mild thermal annealing program consisting of an initial 1-hour incubation at 37 °C, then a cooling step from 37 °C to 22 °C over 2.5 minutes, followed by incubation on 22 °C for 14 hours and finishing with a cooling step from 22 °C to 4 °C. Excess of functional groups were removed using different approaches. Structures functionalized with ephrin-A5-Fc used in paper I. were purified using size-exclusions using two Sepharose 6B-loaded spin columns, structures functionalized with fluorophores used in paper II. were purified using agarose gel-extraction or using size-exclusion purification with Sephacryl S300HR loaded spin columns, while the structures functionalized with rsFPs were immobilized in flow-chambers without prior purification.

3.5.1.2 Characterization of functionalized DNA origami structures

For assessing the success of the functionalization, the structures were run in a 2% agarose gel in 0.5X TBE buffer (10 mM Mg^{2+}) with ice-water bath cooling in order to detect the appearance of fluorescence (in case of rsFPs and fluorophores) and the decrease in migration speed of the modified structures due to their increased size. Generally, gels were run at 90 V for 3 hours, however in the case of ephrin-A5-Fc modified structures, due the small size of the modification, 70 V for 4 hours was used in order to resolve the size difference between modified and unmodified structures.

The functional characterization of structures was performed using different approaches. For ephrin-A5-Fc modified structures used in paper I. agarose gel shift assay was used initially to estimate the functionalization yield of structures with different ephrin-A5-Fc configurations by comparing the band intensities of fractions with detectably different migration speed (due to different functionalization state). Additionally Rhodamine labelled ephrin-A5-Fc was also used to confirm the functionalization. The functionality of the ephrin-A5-Fc carrying structures were assessed in a number of ways. To assess the ability of the ephrin-A5-Fc molecules on the nanostructures to bind Eph-receptor apparent dissociation constants of these structures were measured to the immobilized, extracellular domain of Eph-receptor using surface plasmon

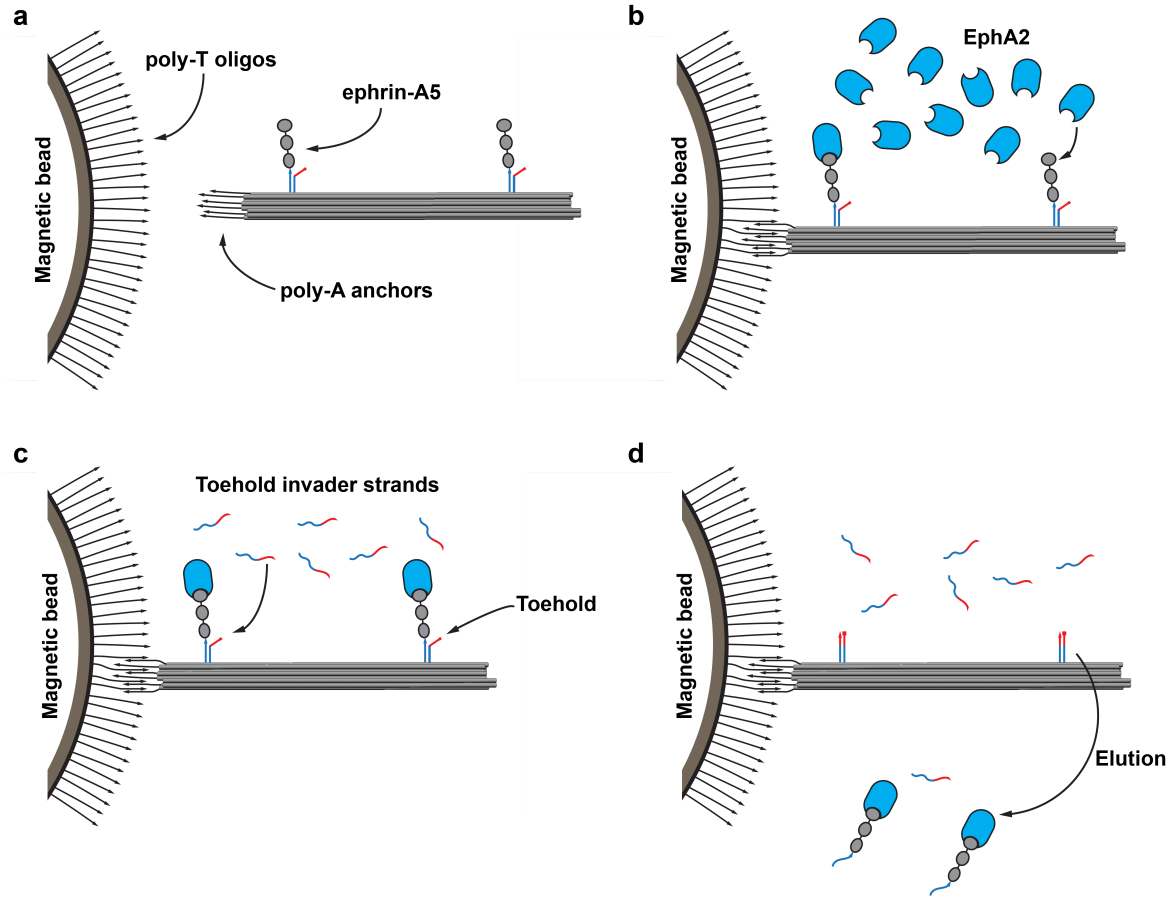

Figure 12. Pull-down assay for characterizing binding capacity of ephrin-A5 functionalized nanostructures. (a) Ehprin-A5 modified nanostructures were bound to poly-T magnetic-beads using poly-A anchors **(b)** structures then were incubated with EphA2-receptor fragments **(c)** and the formed complexes were incubated after washing with invader–oligo, fully complementary to the ephrin-A5-Fc anchoring sites on the structures, for competing the formed ephrin-A5-Fc conjugate-EphA2 complexes off via toehold mediated displacement **(d)** finally the eluents were collected and analyzed with a silver-stained SDS-PAGE gel and western blot.

resonance (SPR). To confirm the binding stoichiometry of the different ephrin-A5-Fc functionalized nanostructures we also performed a magnetic bead–based EphA2 pull-down assay with the ephrin-A5 nanostructures (Fig. 12) for which we made version of the nanorods carrying poly-A anchoring oligos, for immobilization on poly-T functionalized magnetic beads, and ephrin-A5-Fc anchoring sites with extra, 10nt long toehold sequence. We immobilized these structures on magnetic beads and incubated them with EphA2 receptor fragments for 3 hours at room temperature. The receptor/ligand complexes formed on the nanorods were eluted after washing with toehold invader strands with full complementarity to the ephrin-A5-Fc anchoring sites on the structures. The eluates were run in SDS-PAGE gels and were analyzed using silver-staining and western blot.

The functionalization yield of fluorophore-functionalized nanorods used in paper II. was measured using a combined UV-absorbance and fluorescence intensity measurement. The concentration of DNA nanorods was determined by measuring the UV absorbance of the samples at 260 nm, while the concentration of the Alexa Fluor 488-oligos attached to the nanorods was determined by measuring the fluorescence emission (Ex: 489 nm, Em: 519 nm) of the nanorods and converting the obtained fluorescence-intensity values to Alexa Fluor 488

concentration values with a standard curve of Alexa Fluor 488-oligo concentration using a BioTEK SynergyMx Plate-reader.

Finally, the rsFP-functionalized structures used in paper III. were characterized in two different assays. The polygonal, wireframe structures were purified using Sepharose 6B-loaded spin columns, as described before, and the purified, rsFP-functionalized structures were run in EtBr-stained, 2% agarose gel for the determination of origami concentration from band intensities. For measuring the rsFP concentration same volume of the purified structures were run in a 12% SDS-PAGE along a dilution series of rsFP-conjugates. The gel was silver stained and the rsFP concentration of the samples was then calculated from the protein band intensity values using the calibration curve produced with the reference samples. For measuring the functionalization yield of the nanorods used in paper III. a modified version of the earlier described bead immobilization assay was used. Unpurified, rsFP functionalized nanorods were immobilized on Streptavidin functionalized magnetic beads via their biotin tags used for surface immobilization. After the unbound rsFP-conjugates were washed away, the rsFP conjugates bound to the structure were eluted using warm SDS-PAGE loading buffer. Protein concentration was determined as described for the sheet structures. Origami concentration was determined from the band intensities of the unpurified structures used for bead-immobilization corrected by the band intensities of the unbound and washing fractions of the bead experiment in an EtBr-stained, 2% agarose gel.

3.6 Super resolution imaging of DNA origami structures

3.6.1 Sample preparation for imaging applications

Two strategies were used in the work presented in this thesis for immobilizing DNA origami structures on microscope coverslips for imaging experiments. Fluorophore-functionalized nanostructures used in paper II. were immobilized using electrostatic interaction. The Alexa Fluor 488-labeled DNA origami samples were diluted to 35 pM with storage buffer (5 mM Tris, 1 mM EDTA, 10 mM $MgCl_2$) containing DABCO as antifading agent, the CAGE 552-labeled DNA nanostructures were diluted to 35 pM with standard storage buffer. Coverslips were pretreated with glow discharging to introduce negative charge to the surface, which was used together with the Mg^{2+} in the buffer to bind the inherently negatively-charged nanostructures to the surface. Diluted DNA origami samples were spotted on glow discharged coverslips and were incubated for 10 minutes in the dark at room temperature, then coverslips were washed with storage buffer blotted off and inverted on a microscopy slide and sealed.

Structures imaged with DNA-PAINT imaging presented in paper III. and IV. were immobilized using biotin anchors incorporated in the structures by binding them to a biotinylated-BSA/Streptavidin pretreated surface. For one-color DNA-PAINT imaging performed in paper III. flow chambers were built by sticking two strips of double-sided scotch tape ~8 mm apart on microscope slides and sticking a coverslip on top. Solutions were washed through the assembled flow chambers by pipetting buffer from one side and sucking liquid out on the other side with lab wipes. The flow chamber was incubated first with 1mg/mL

biotinylated-BSA in Buffer A (10 mM Tris, 100 mM NaCl, 0.05% Tween-20, pH 7.5) for 2 minutes. After washing with Buffer A the chamber is incubated with 0.5 mg/mL of Streptavidin in Buffer A for 2 minutes. The chamber was then washed with Buffer A and Buffer B (5 mM TRIS, 10 mM $MgCl_2$, 1 mM EDTA, 0.05% Tween-20, pH 8.0) before polyhedral sheet and nanorod structures diluted in Buffer B to 50 pM or 100 pM respectively were incubated in the chamber for 5 minutes. After washing the chamber with Buffer B the imager solution had been washed in containing the Atto 550-labelled imager strands (1.7 pM-10 pM Atto 590-P2), an oxygen scavenging system (2.5 mM protocatechuic acid (PCA) and 7 µg/mL Protocatechuate 3,4-Dioxygenase (PCD)) counteracting bleaching and a triplet-quencher (1 mM TROLOX), for preventing blinking of the fluorophores, in Buffer B and the chambers were sealed with two component epoxy glue. For multi-color, Exchange-PAINT experiments performed in paper IV. glass bottom ibidi (µ-Slide VI 0.5) channel slides were used. The sample preparation was nearly identical to assembled flow chambers with longer incubation times (5 minute incubation for biotinylated-BSA/Streptavidin and 10 minute incubation for the structures) and an extra step introduced before the immobilization of the structures for flowing in 80 nm gold nanoparticles, used for drift correction and channel alignment, in Buffer A and incubating them for 5 minutes before washing the channel. Each imaging round was performed using an imager solution containing a particular DNA-PAINT imager strand (10 nM Atto 590-P1, P2, P4 or P6, oxygen scavenging system and TROLOX). Channels were washed 6-times with Buffer B in between rounds to remove the current imager solution before the introduction of the new one.

3.6.2 One and multi-color DNA-PAINT imaging

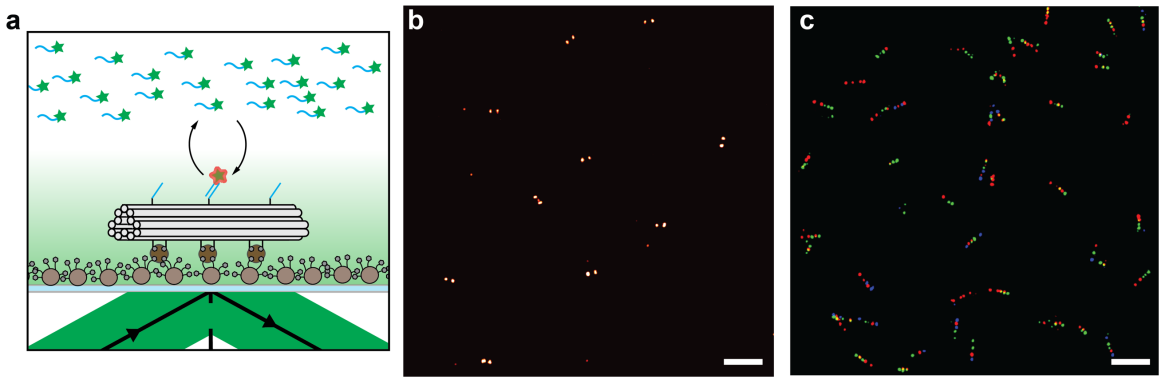

Figure 12. Super resolution DNA-PAINT imaging of DNA-origami structures. (a) Illustration of the working principle of DNA-PAINT super resolution imaging used in this thesis: DNA-PAINT docking site containing DNA-origami structures were immobilized in streptavidin coated flow-chambers using biotin anchors. Blinking events for super resolution imaging were generated through the detection of the transient binding events of short, fluorophore-labelled imager-strands to the docking sites on the structures over the background of the free-floating imager-strand pool using TIRF illumination. **(b)** One-color DNA-PAINT super resolution image of nanorods used in paper III. with the imaged rsFP-anchoring sites at 100 nm from each other (scale bar = 500 nm). **(c)** Three-color Exchange-PAINT super resolution image of DNA-origami barcodes (v1) used from paper IV. (scale bar = 500 nm).

One-color DNA-PAINT and multicolor Exchange-PAINT imaging of nanostructures was performed in paper III. and paper IV. respectively (Fig. 12). DNA-PAINT is an SMLM super resolution technique, where the blinking events are generated by 9 bp long Atto 550-labelled DNA-PAINT imager sequences transiently binding to their complementary docking sequences displayed on the structures. For all experiments a Nikon Eclipse Ti-E microscope with the

Perfect Focus system (eliminating Z-directional drift during long acquisitions) and an objective-type TIRF configuration using an iLAS2 circular TIRF module was used. The TIRF illumination was used to restrict the excitation to a calculated penetration depth of ~90-180 nm to decrease the background fluorescence generated by the pool of imager strands present in the solution. A 1.49 NA CFI Plan Apo TIRF 100× Oil immersion objective was used for magnification for all experiments with an extra 1.5X magnification in some cases achieving a final pixel size of 130 nm and 87 nm respectively. For illumination an OBIS 561 nm LS 150mW laser was used with custom iLas input beam expansion optics (Cairn) optimized for reduced field super resolution imaging with corresponding clean-up and emission filters. An iXon Ultra 888 EMCCD camera (Andor) was used for recording with a 512 pixel by 512 pixel ROI centered on the illuminated area. Time lapses of experiments with the length of 6000 to 9000 frames were acquired with micromanager software using frame-transfer mode of the camera, 300 msec exposure time, 10 MHz readout rate and no EM gain. Localizations were detected in the recorded movies using the Picasso software platform developed by Jungmann and co-workers[90] using the MLE algorithm and stage drift correction using the redundant cross-correlation (RCC) algorithm. A consecutive filtering step was performed using the Picasso Filter software based on the uniform event shape and uniform, high localization precision to remove low quality and multi-localizations. Reconstructed data was exported as .hdf files containing lists of localization coordinates and super resolution images using 20X sub-pixelation.

For multicolor DNA-PAINT imaging performed in paper IV. an approach called Exchange-PAINT was used, where DNA-PAINT docking sites and imager strand pairs with different sequence are used in consecutive rounds to create different colors. This allowed the use of imager strands with the same fluorescent group (Atto 550), making it possible to align channels without any correction for chromatic aberration. The alignment of the channels was done with the help of 80nm gold nanoparticles included in the sample preparation, using the Picasso Render software's *Align channels* feature utilizing a cross-correlation based algorithm.

3.6.3 Super resolution image processing

3.6.3.1 Detection of structures in DNA-PAINT images and processing localization within them

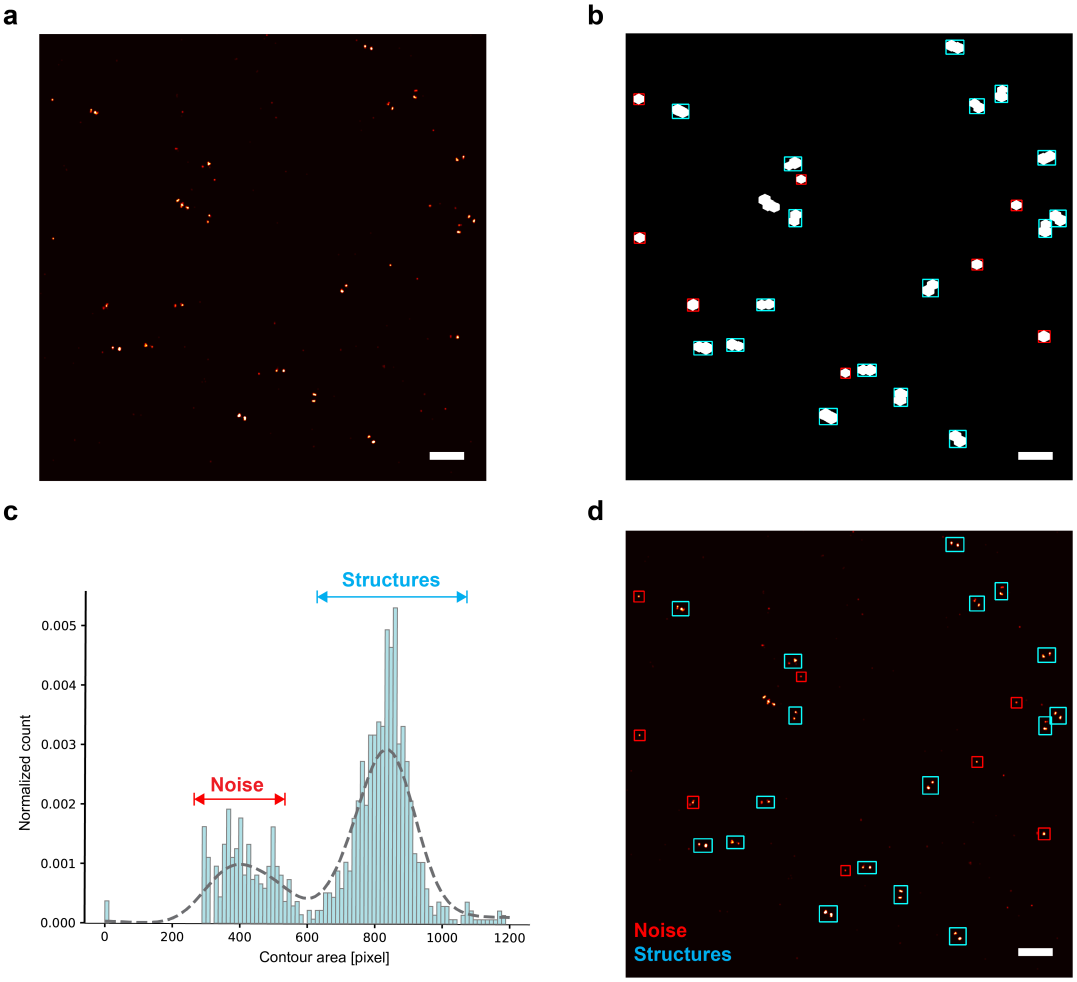

Figure 13. Detection of DNA-origami structures in DNA-PAINT images. Workflow for the DNA origami structure detection in DNA-PAINT images represented on the images of 100nm DNA origami rods used in paper III.: **(a)** Reconstructed super resolution (SR) DNA-PAINT images of DNA origami nanorods (scale bar = 500 nm). **(b)** Same image after grayscale conversion, filtering with Otsu's binarization and pixel dilation with detected contours clustered into groups based on their area (blue and red ROIs) (scale bar = 500 nm). **(c)** Normalized contour area distribution plot with the noise and structure contours' size range determined from the peaks detected in the kde estimation of the distribution (gray dashed curve). **(d)** Original DNA-PAINT image of the DNA origami nanorods with the identified structures (cyan ROIs) and noise (red ROIs) highlighted (scale bar = 500 nm).

All image processing performed on DNA-PAINT images of DNA nanostructures presented in paper III. and paper IV. were done by custom-written Python scripts using the Open Source Computer Vision Library. A common feature of the these were the detection and position determination of nanoconstructs in SR microscopy images (Fig. 13). As a general approach we first transformed the images to grayscale and performed an automatic thresholded binarization using Otsu's method. For the enhancement of the images of the structures in the binary image we utilized a pixel dilation step. We then detected the contours of objects in the images and could identify the contours belonging to DNA nanostructures, based on the contour area. We were then able to determine the position of structures by calculating the center position of the contour's minimal enclosing circle.

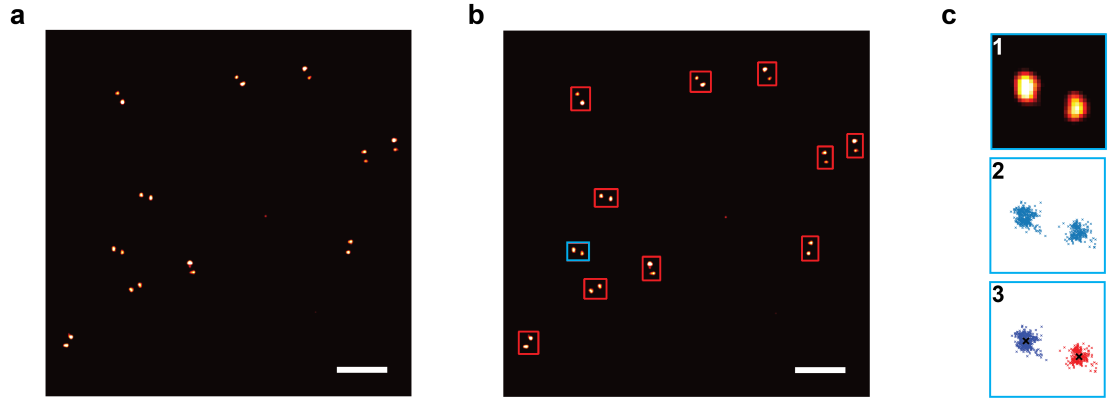

Figure 14. Using DNA-PAINT localization data to characterize DNA-origami structures. Workflow for the anchoring site distance determination from DNA-PAINT images represented on the images of 100nm DNA origami rods used in paper III.: **(a)** Reconstructed super resolution (SR) DNA-PAINT images of DNA origami nanorods (scale bar = 500 nm). **(b)** DNA origami nanorods detected in the images (red ROIs) with earlier described contour-detection approach. **(c/1)** Image of DNA origami nanorod cropped out of the SR image (cyan ROI) using the determined ROI coordinates. **(c/2)** Plot containing the coordinates of the localizations in the ROI, exported from the SR image's localization file. **(c/3)** Plot containing the coordinates of the localizations in the ROI sorted into two groups (blue/red) using k-means clustering and the mean positions of the two clusters (black cross), that were used to calculate the distance between the centroid of the sites (black cross).

In the work presented in paper III. we used DNA-PAINT data to characterize the DNA nanorods in terms of the distance distribution and functionalization state of their rsFP anchoring sites. For measuring the distance distribution, we functionalized the structures with complementary oligonucleotides to the rsFP anchoring sites containing a DNA-PAINT docking site. After imaging the structures following the procedure described earlier, we determined the coordinates of ROIs in the image containing single origami nanorods using the contour detection approach detailed in the section before. Using these coordinates we then were able to extract the coordinates of the localizations belonging to the structures from the original DNA-PAINT localization file. We then grouped the localizations into two cluster using k-means clustering and determined the coordinates of the sites by calculating the average of the localization coordinates belonging to the two clusters. We finally calculated the distance of the sites from the determined site coordinates (Fig. 14).

We also used DNA-PAINT to assess the functional state of the sites on rsFP fuctionalized nanorods. For this we prepared rsFPs conjugates with anchoring oligonucleotides containing a DNA-PAINT docking site. By functionalizing the nanorods with these conjugates we were able to directly image the proteins on the DNA origami nanorods using DNA-PAINT. As the blinking events are generated by the binding events of imager strands to docking sites, the frequency of events scales with the number of docking sites or docking site labelled molecules in a ROI[83]. By detecting individual rsFP-labelled DNA nanorods in the generated SR image, we were able to extract the localizations belonging to the particular DNA nanorods and assign them to the two rsFP functionalization sites on the nanorods using the approach described before. We were then able to compare the functionalization state of the rsFP sites on the nanorods between and within structures.

3.6.3.2 Barcode identification from multicolor DNA-PAINT images

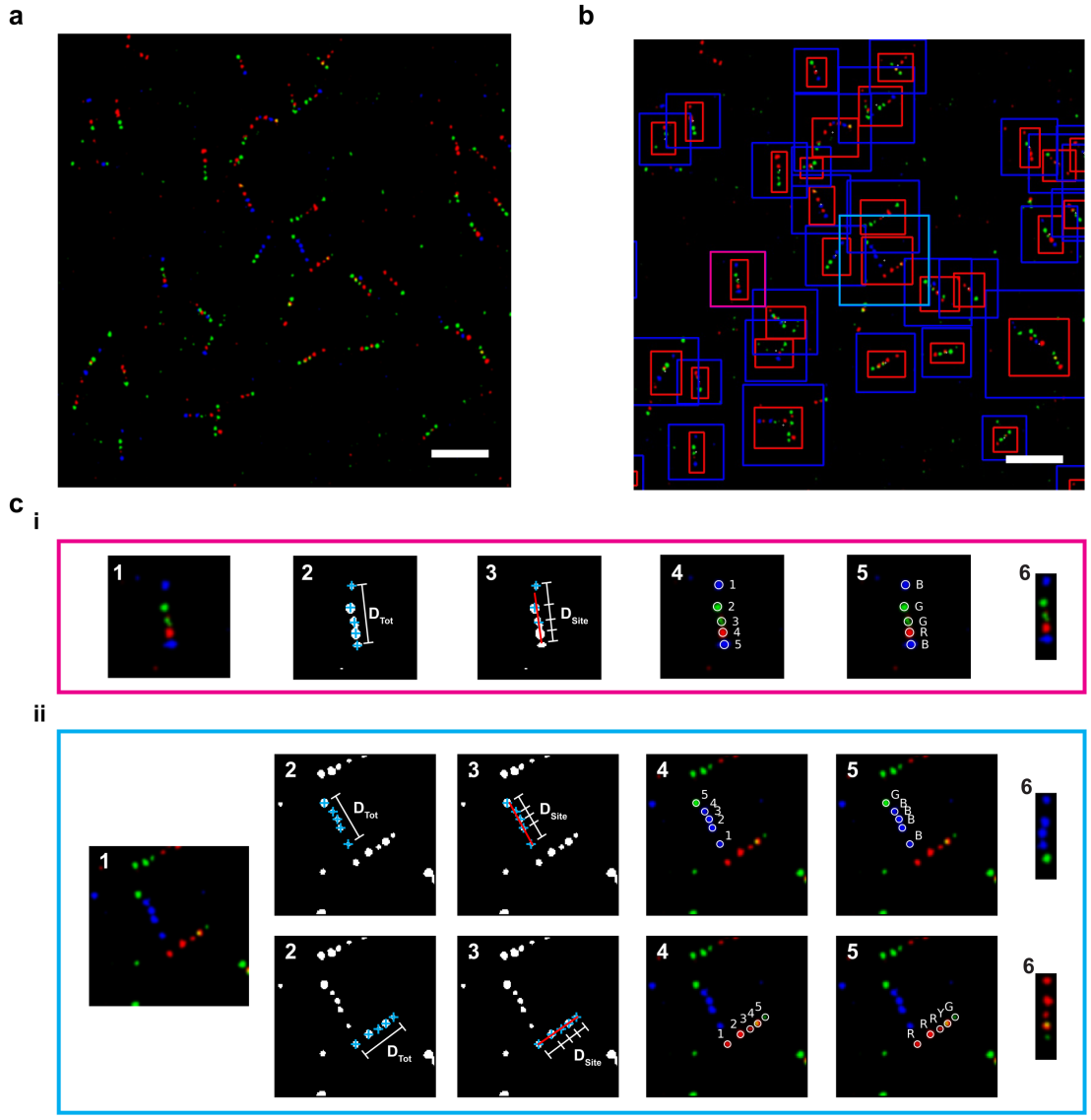

Figure 15. DNA-origami barcode identification in DNA-PAINT images (a) Three-color Exchange-PAINT image of DNA origami barcodes (v1) immobilized in glass bottom, flow chamber slides (scale bar = 500 nm). **(b)** Same image with detected barcodes highlighted with red, rectangular bounding ROIs, bounding coordinates for the ROIs used for color identification highlighted in blue and ROIs used for demonstrating color determination highlighted in magenta and cyan. **(c/i)** Examples of processing steps performed on single barcode containing ROIs and **(c/ii)** multiple barcode containing ROIs for determining color combination of barcodes in them: **(c/i-ii/1)** ROIs were transformed into **(c/i-ii/2)** binary, thresholded images and all sites' coordinates (blue crosses) were determined by blob detection using the Laplacian of Gaussian (LoG) method. Barcodes were checked for largest relative site distance (D_{Tot}) and linearity (by calculating the linear least-squares regression for the site coordinates). This step was performed for all five site combinations for multiple barcode containing ROIs. **(c/i-ii/3-4)** For barcodes who passed these steps the order of sites was determined by the relative distances of sites (D_{Site}), and a site distance thresholding was performed on the barcodes. **(c/i-ii/5)** For barcodes that passed the thresholding colors of the positions was determined from the RGB values within the detected position ROIs (white circles). **(c/i-ii/6)** Finally barcode images were vertically aligned and cropped.

Multicolor Exchange-PAINT images were generated of DNA origami barcode library version 1 and 2. Barcode structures were detected in SR microscopy images and their coordinates determined using the approach described before. Images of the barcode containing ROIs were binarized and barcode sites in the image were detected with scikit-image package's blob detection feature using the Laplacian of Gaussian (LoG) method. ROIs with incomplete barcodes or noise (less than 3 sites detected) were discarded, images with multiple barcodes

(more than 5 sites detected) or a single barcode (five sites detected) were further processed. For images with one barcode the linearity of the sites was checked by calculating the linear least-squares regression for the site coordinates and the maximum relative distances of the sites was also filtered for. For barcodes satisfying these criteria, the site positions were determined using the relative distances of sites and the color of each site was determined based on the average RGB values in each site. For images with multiple barcodes each combination of five sites was tested for linearity and relative distances and non-overlapping sets of five sites satisfying these criteria were kept. The color combination of each barcode was determined as for single barcode images (Fig. 15).

3.7 Next generation sequencing of DNA origami barcodes

The color encoding inserts of DNA origami barcodes used in paper IV. were characterized with NGS before and after cloning to assess the diversity and homogeneity of the constructed libraries. For barcode library version 1 PacBio RS II system was used for sequencing the insert after assembly and after cloning, due to its long permitted read length. For barcode library version 2, because of the decreased size of the insert, we were able to use 2x300 bp paired-end sequencing on the Illumina MiSeq platform. All samples were prepared using PCR of the barcode insert not containing the KAN_R gene followed by purification using agarose gel extraction. The sequencing sample preparation and sequencing were performed by GATC biotech for barcode library version 1 and by the Swedish National Genomic Infrastructure (NGI), Stockholm for barcode library version 2. All NGS data was processed with custom Python scripts using the Biopython library (Fig. 16). An initial quality control was performed on all reads by excluding shorter reads. Following that pairwise alignments were performed between all possible color-site sequences and the color-site subsequences in the reads (with a +30 nt window for library version 1 and +5 nt for library version 2 on each side to account to deletions and insertions along the read) using a gap opening and extension penalty of -1 and a match a mismatch score of +1 and -1 respectively and no terminal gap penalty. For each color reference sequence, the best alignment was used to calculate a confidence score for that particular color by weighting the alignment score of each position in the alignment by the Phred score corresponding to the read base in the alignment. The color site sequence with the highest identity score was declared as the winner and the site color was called. In the case of winner color sequences that are unacceptably similar to the second candidate (<5% score difference between winner and second candidate) or have high number of mismatches to the read (<30% identity) the site was called as an ambiguous site. The results were exported as a Barcode ID file containing the called color-combination of the insert, its compound identity score and alignment information for all sites (sequence, called color, identity score for called color site

sequence, second highest color, identity score for second highest color site sequence) and the UMI sequence and its Phred score for library version 2.

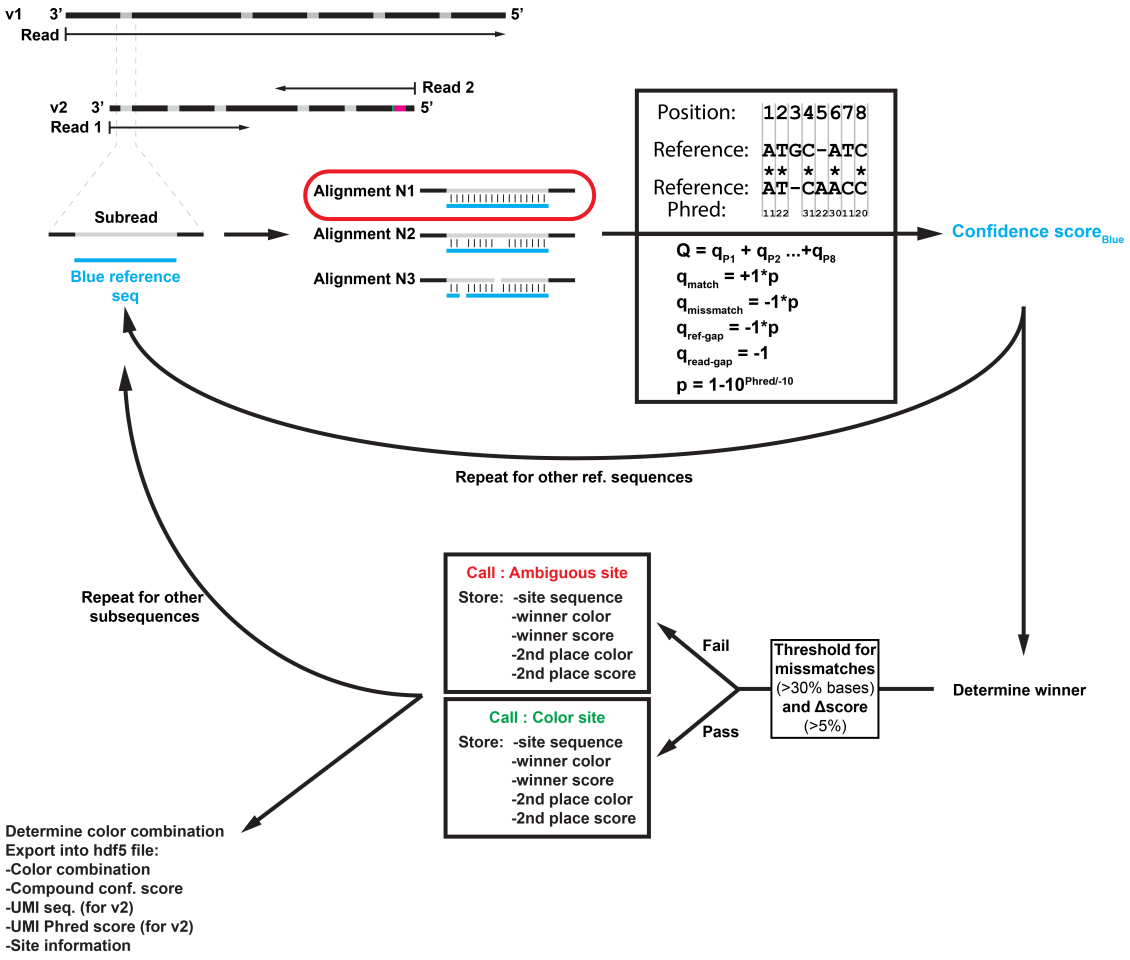

Figure 16. Workflow of color combination identification from NGS data

For barcode library version 2 an additional set of scripts were used to create the barcode-UMI catalogue file. First the Barcode ID file was parsed through and the UMI sequences along their Phred scores for each color combination were sorted into separate MULTI-FASTA files with low confidence bases (Phred<30) in the UMI read sequences distinguished by lowercase letters. Similar UMI-sequences were grouped together using the CD-HIT[94] clustering algorithm and a consensus sequence was generated for each cluster using the ClustalW2[95] program. The UMI consensus sequences and their cluster size were finally exported along their linked barcode color combinations and their confidence scores into a barcode-UMI catalogue.hdf5 file.

4 RESULTS AND DISCUSSION

4.1 PAPER I.

4.1.1 Production and characterization of ephrin functionalized DNA origami nanorods

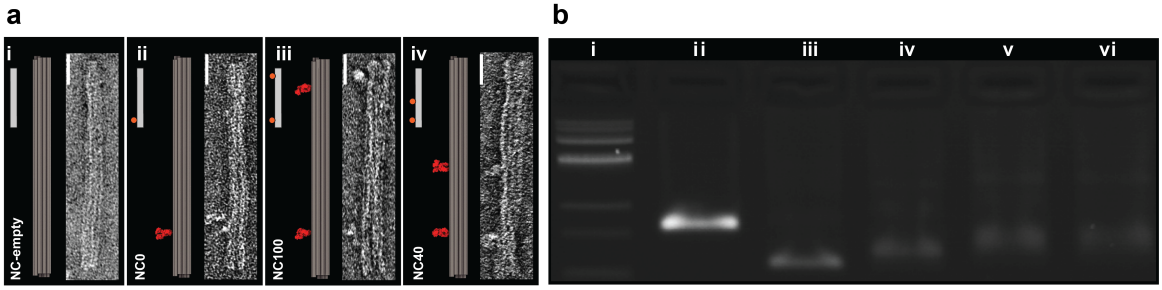

Figure 17. Ephrin A5 functionalized DNA origami nanorods (a) Representative electronmicrographs of DNA origami nanorods functionalized with **(a/i)** zero (NC-empty), **(a/ii)** one (NC0), two ephrins positioned **(a/iii)** 101.1 nm (NC100) and **(a/iv)** 42.9 nm (NC40) from each other with ephrin-A5-Fc molecules visibly positioned at predesigned patterns (scale bars = 20nm). **(b)** Nanorods with zero (iii), one (iii) and two ephrin-A5-Fc molecules positioned at 101.1 nm (v) and 42.3 nm (vi) run along the p7560 scaffold (ii) and DNA ladder in an EtBr-stained 2% agarose gel showing decrease in migration speed with the increasing number of ephrin-A5 molecules.

In paper I. we utilized DNA origami to study the effect of receptor clustering in the Eph-receptor/ephrin signaling pathway. For this we constructed a set of DNA origami nanorods, also called as nanocalipers (NC), carrying functionalization sites for positioning ephrin-A5 ligands at different distances, at 101.1 nm (NC100) and 42.9 nm (NC40), along with structure carrying only one site (NC0) and zero site (NC-empty) for control experiments. After we produced and purified the structures following the procedure described earlier we used the anchoring oligonucleotide-ephrin-A5-Fc conjugates, produced with the previously described hydrazide-hyrazone click chemistry, to functionalize the structures. For confirming the functionalization, the nanocalipers were imaged using TEM and tested in an agarose gel shift assay, both of which showed successful attachment of the ephrin-A5-Fc-olignucleotide conjugates (Fig. 17) with high yield (close to 90 %). For testing the binding capacity of the ephrin-A5-Fc-decorated nanocalipers a number of different methods were utilized. Binding affinity of the oligonucleotide conjugated ephrin-A5-Fc alone or positioned on nanocalipers (NC0, NC40 and NC100) was compared to unconjugated ephrin-A5-Fc using SPR measurements of the apparent dissociation constants of these samples to the extracellular domain of EphA2 receptor, which showed a similar affinity of the conjugate and NC0 to the unconjugated ligand and an increased affinity in the case of NC40 and NC100, due to the higher number of ephrins positioned on them. We also measured the binding capacity of the nanocalipers to the EphA2 receptor using a pull-down assay, described in detail earlier, by measuring the amount of bound EphA2 receptor fragments to magnetic bead-immobilized nanocalipers. This assay has shown a clear trend for structures carrying increasing amount of ephrins (one, two and three) and confirmed the nearly identical binding capacity of the structures displaying ligands at different distances (~40 nm, ~80 nm and ~ 100 nm) (Fig. 18).

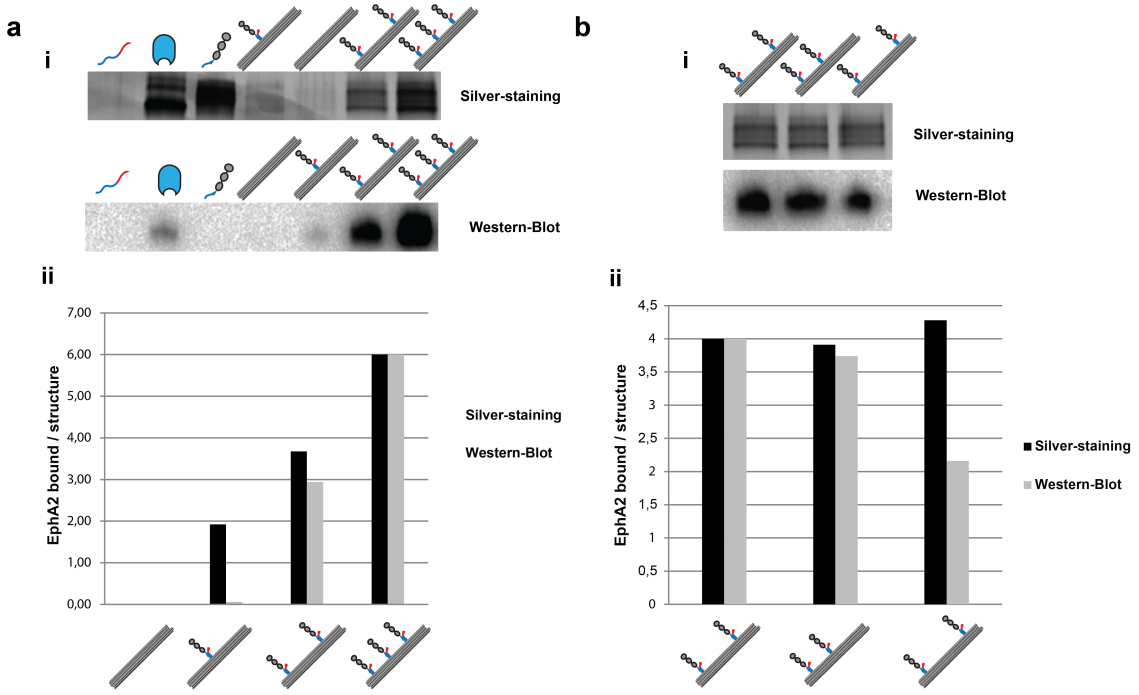

Figure 18. Measuring the EphA5 binding capacity of ephrin nanocalipers using a pull-down assay (a/i) Gel images of eluates from pull-down assay performed with nanocalipers with zero, one, two and three ephrins run in 15 % denaturing PAGE and detected with silver-staining (top) and western blot (bottom) **(a/ii)** along with a graph displaying band-intensity values normalized to the three-site nanorod, showing a clear increase of binding capacity with more ephrin-ligands displayed on the nanorods. **(b/i)** Gel-images of eluates from pull-down assay performed with nanocalipers with two ephrins at ~40 nm (NC40), 80nm (two-site nanocaliper) and ~100 nm (NC100) from each other run in 15 % denaturing PAGE and detected with silver-staining (top) and western blot (bottom) **(b/ii)** along with a graph displaying band-intensity values normalized to the two-site nanorod, showing a nearly identical binding capacity of nanorods displaying two Eph-receptor ligands at different distances.

4.1.2 Stimulation of cancer cells with ephrin functionalized DNA origami nanorods

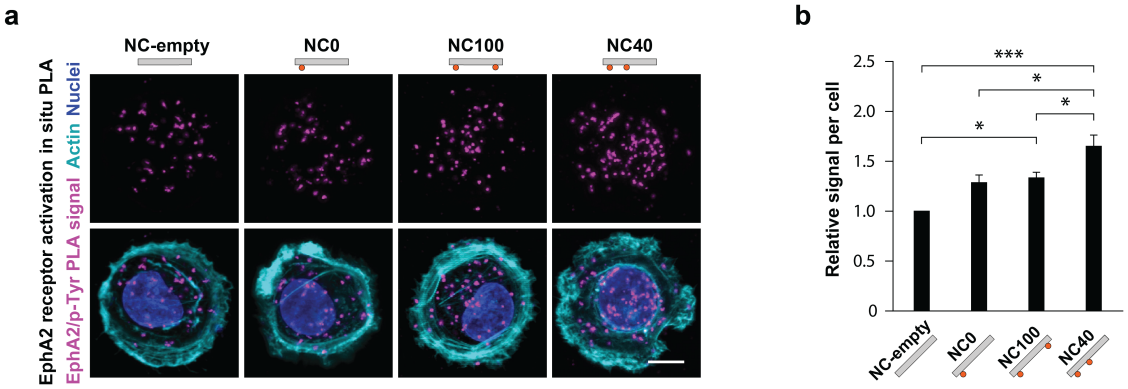

Figure 19. Measuring EphA2 receptor activation in MDA-MB 231 cells stimulated with ephrin-nanocalipers (a) Fluorescent images of MDA-MB 231 breast cancer cells stimulated with nanocalipers displaying zero (NC-empty) one (NC0) and two ephrin-ligands at displayed 100 nm (NC100) and 40 nm (NC40) from each other, with EphA2 receptor phosphorylation detected with PLA (magenta), nuclei stained with DAPI (blue) and actin cytoskeleton stained using Alexa Fluor 488–phalloidin (turquoise) (scale bar = 10 µm). **(b)** Bar graph comparing the normalized, average number of activated EphA2 receptors per cell upon the treatment with different ephrin-nanocalipers, showing a significantly highest level of activation in the case of NC40 (*P<0.05, **P<0.001).

We measured the effect of nanostructures presenting ephrinA5-molecules at different distances on the EphA2-receptor signaling by incubating ligand-concentration normalized amounts of nanocalipers, displaying ephrinA5 molecules at different distances, with MDA-MB 231 breast

cancer cells cultured on glass slides micropatterned with fibronectin islands (in order to avoid "self-activation" through cell-cell contacts) and detecting EphA2 receptor activation (more accurately the phosphorylation of Tyrosine residue in the intracellular domain of the receptor) with an in situ Proximity Ligation Assay (PLA) (Fig. 19). We observed that the calipers that displayed ephrinA5 ligands (NC0, NC40 and NC100) induced higher level of activation than the empty nanocalipers (NC-empty) alone. Additionally, we saw that the stimulation with NC40 led to the highest receptor phosphorylation, suggesting a ligand proximity induced mechanism of Eph receptor activation. To further strengthen this finding, we also studied the effect of nanostructures presenting EphrinA5-molecules at different distances on downstream cellular-signaling events with a Cell Invasion Assay. The migration of MDA-MB 231 breast cancer cells, pre-stimulated with different ephrin-nanocalipers through extracellular matrix coated cell culture inserts was quantified, and in agreement with the results from the receptor activation experiment the cells stimulated with NC40 structures showed the greatest decrease in invasiveness that is understood to be induced by EphA2 receptor activation[96].

4.2 PAPER II.

4.2.1 Production and characterization of fluorophore functionalized DNA origami nanorods

In paper II. we used DNA origami structures as reference samples for estimating the localization accuracy of fluorophore positions for two commonly used SR microscopy methods, STORM and STED. For this purpose, we constructed rigid, DNA origami nanorods with twelve fluorophore functionalization sites approximately 100 nm (105.5 nm) from each other (Fig. 20). This structure was then used to position fluorophores (Alexa Fluor 488 and CAGE 522) in a resolvable, uniform and conserved distance. As the structure was used as reference sample it was crucial to understand the variance of the distance originating from the sample. We identified two main sources contributing to this: the distortion of the structures and the incomplete functionalization of the nanorods. The former we measured by calculating the longitudinal length distribution of nanorods from TEM micrographs, where we measured the mean length to be 141.6 ± 4.1 nm, compared to the designed length of 140.4 nm, which resulted in a variance of 16.6 nm^2 (Fig. 23). The incorporation efficiency of fluorophores was measured for a purified Alexa Fluor 488-functionalized nanorods by measuring the nanorod concentration by UV absorption at 260 nm and the Alexa Fluor 488 concentration from the 488nm fluorescence of the structures using an Alexa Fluor 488-oligonucleotide calibration curve. The occupancy of the nanorods was measured to be 96.0±2.4 % resulting in a 0.05 nm^2 distance variance based on the designed positions of the 24 sites. As the compound

variance was measured to be much lower than the sensitivity of the two imaging techniques we deemed the structures to be suitable reference samples for accuracy estimation.

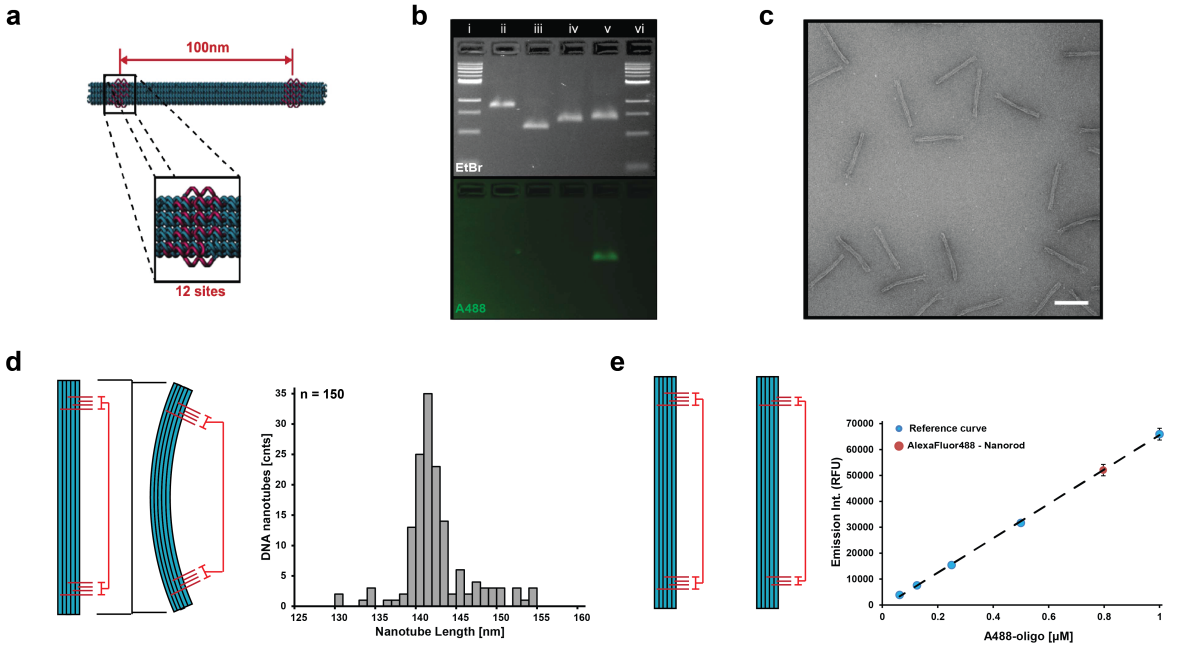

Figure 20. DNA origami reference samples for STED and STORM imaging (a) Computer rendering of the DNA-origami nanorod with 12-12 addressable labeling-sites (scarlet) at approximately 100nm distance. **(b)** Alexa Fluor 488-labelled DNA origami 100 nm nanorod (v) run along unfunctionalized 100 nm nanorod (iv), empty nanorod (iii), scaffold ssDNA and DNA ladders (i, vi) in an EtBr stained 2% agarose gel. **(c)** Transmission electron microscopy micrograph of origami nanorods (scale bar = 100 nm). **(d)** The structural distortion approximated by the longitudinal length distribution of nanorods measured from TEM micrographs. **(e)** Measuring fluorophore attachment yield by calibrated fluorescence measurement of Alexa Fluor 488-labelled nanorods.

4.2.2 Testing STORM and STED microscopy's accuracy with DNA origami nanorods

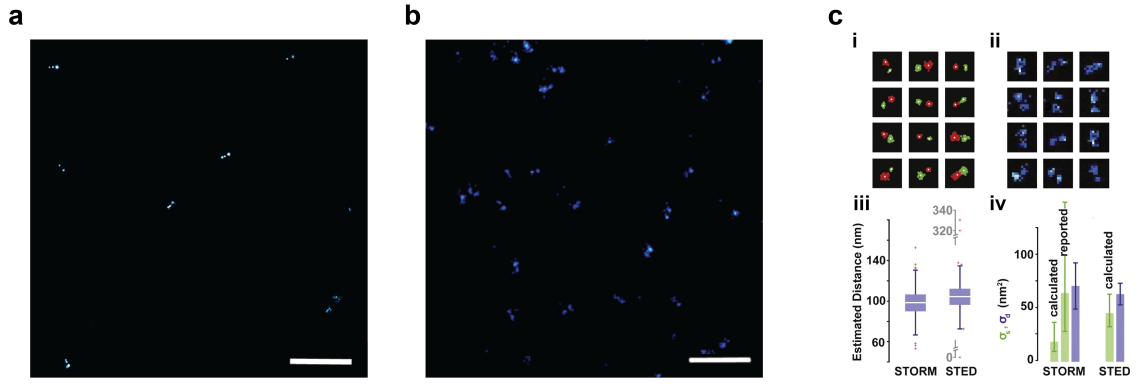

Figure 21. Measuring accuracy of STED and STORM imaging with DNA origami nanorods (a) Field of view images of DNA origami nanorods imaged with STORM **(b)** and STED (scale bar = 1 µm). **(c)** Representative images of nanorods imaged with **(c/i)** STORM and **(c/ii)** STED. **(c/iii)** Comparison of distance distributions for STORM (left) and STED (right). **(c/iv)** Measured position accuracy (purple) and precision (green) for STORM and STED.

Classically the metric used for evaluating the performance of an imaging technique is the variance of estimated position of a molecule. We argued that, as this is more of a measure for the reproducibility of the measurement of the molecule's position, this metric does not inform us about the actual position of the molecule imaged, thus it can be blind to systematic errors.

We instead derived an estimator for positional accuracy. As the relative, true positions of fluorophores are known in the nanorods, one can use the two positions on the nanorods as each other's reference. We calculated this measure of accuracy from the variance of distance between the two sites, as it is the sum of the positional variances of the individual spots, because the sites' covariance is negligible at resolvable distances. We first tested the performance of our accuracy measure in simulations of emitter pairs. The results showed that our metric performs better at shorter emitter distances and actually picks up positional bias compared to the commonly used localization precision measure. Following this, we acquired super resolution images of the fluorophore-functionalized nanorods electrostatically immobilized on coverslips using STED- and STORM-imaging (Fig. 21). The images were processed and accuracy and precision values were calculated using a custom written MatLab software, showing that in spite of STORM's higher precision STED had higher accuracy and also estimated the measured distance closer to the designed one (Fig. 21).

4.3 PAPER III.

4.3.1 Production and characterization of rsFP functionalized DNA origami nanosheets and nanorods

The development of super resolution approaches, like RESOLFT, that use reversibly switchable fluorescent proteins (sFPs) has enabled real time *in vivo* imaging of cellular features with sub-diffraction resolution. The achievable spatial resolution and recording speed of these techniques is tightly linked to the photophysical behavior of these proteins. A detailed understanding is lacking about the photo-switching behavior of the rsFPs as our understanding of it is based mostly on bulk measurements, while features in nanoscopy images are closer to the single molecule regime. For this reason, we developed an DNA origami platform to study the photophysical properties of rsFP molecules in a quantitative fashion by controlling the number and the relative position of these rsFPs in microscopy samples for RESOLFT imaging. For studying quantitative dependence of the photophysical parameters of two commonly used rsFPs (rsEGFP2 and rsEGFP(N205S)) we designed a 118nm by 82nm wireframe, DNA origami nanosheet (PGS). We designed the sheets to have 10 (PGS 10S), 20 (PGS 20S) or 40 rsFP anchoring sites (Fig. 22) along with Atto 590 markers, DNA-PAINT[97] docking sites and biotinylated anchoring oligos for surface immobilization (Fig. 22). For testing the resolvable distance with RESOLFT, we designed DNA origami nanorods with 18-18 (18S) or 9-9 rsFP anchoring sites (9S) at ~60 nm, ~80 nm and ~100 nm from each other. These structures also were designed to have Atto 590 markers and biotin surface immobilization sites (Fig. 23). The structures' stability in the buffers used in RESOLFT experiments were tested using DNA-PAINT imaging: in the case of the PGS structures no substantial effect was observed on the

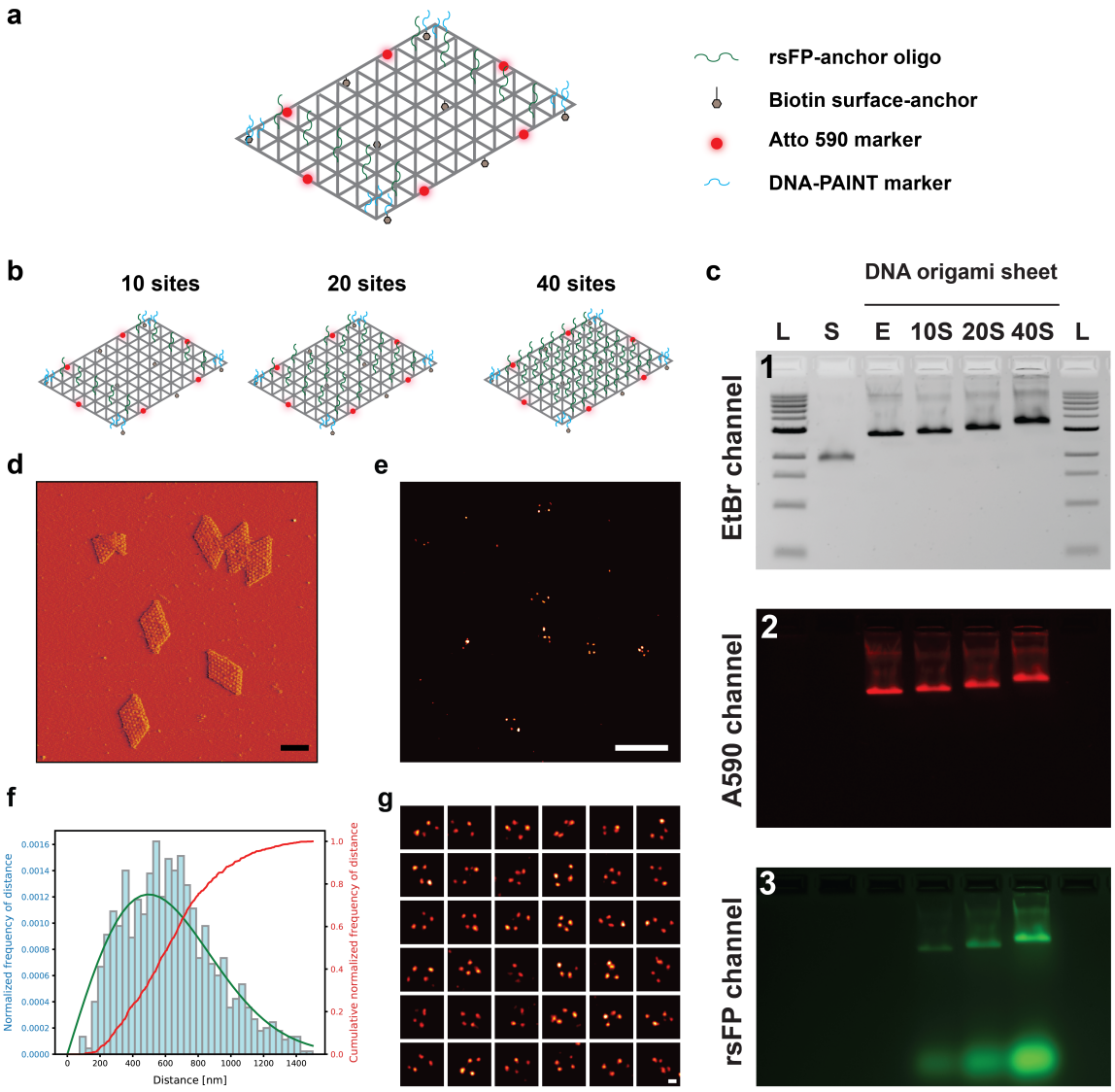

Figure 22. Polygonal, wireframe DNA origami nanosheet used for rsFP characterization (a) Schematic representation of the ~118 nm by ~82 nm triangulated polygonal DNA-origami nanosheet (PGS) used in the experiments with edges representing double-helices, highlighting the different functional groups on the structure. **(b)** Schematic representation of the different versions of the DNA origami nanosheets used in experiments with 10, 20 or 40 rsFP anchoring sites. **(c)** rsFP functionalized DNA origami nanosheets with 10 (10S), 20 (20S) and 40 sites (40S) along with p8634 ssDNA scaffold (S), empty DNA origami nanosheet (E) and DNA ladders (L) ran in 2% agarose gel stained with ethidium bromide and imaged with UV (1), red light (2) and green light (3) illumination. **(d)** AFM image of the purified empty DNA origami nanosheets (scale bar = 100 nm). **(e)** Field-of-view DNA-PAINT super resolution image of DNA origami nanosheets imaged in RESOLFT Buffer (scale bar= 500 nm). **(f)** Plot of the normalized closest neighbor distance frequencies of PGS immobilized on coverslips as histograms (blue) and as cumulative plot (red) and the nearest neighbor distribution function calculated from the measured structure density (green). **(g)** Close-up DNA-PAINT super resolution images of DNA origami nanosheets imaged in RESOLFT Buffer (scale bar= 50 nm).

appearance of the structures (Fig. 22), while for the nanorods a only a negligible difference was observed in the measured distances of the sites compared to standard buffer conditions with the measured distances matching the designed distances quite closely(18S 60 nm: 57.1 nm/56.2±6.7 nm , 18S 80 nm: 78.5 nm/75.1±6.8 nm, 18S 100 nm: 100.0 nm/94.2±6.4 nm designed/measure distance) (Fig. 23). For the origami sheet structures used for the photophysical characterization experiments we also optimized the surface immobilization density of the structures to result in a mean of 632.9±281.9 nm closest neighbor distance between the structures in order to achieve close to a single origami sheet structure per diffraction limited spot in the final RESOLFT images (Fig. 23). For immobilizing the rsFPs on

the origami structures we prepared rsFP-anchoring oligonucleotide complement conjugates from transgenically produces rsFPs using a site-directed, two-step reaction described earlier, that showed no substantial effect on the photophysical parameters of the rsFP molecule. These rsFP-oligonucleotide conjugates were used for functionalizing the origami structures (Fig. 22-23).

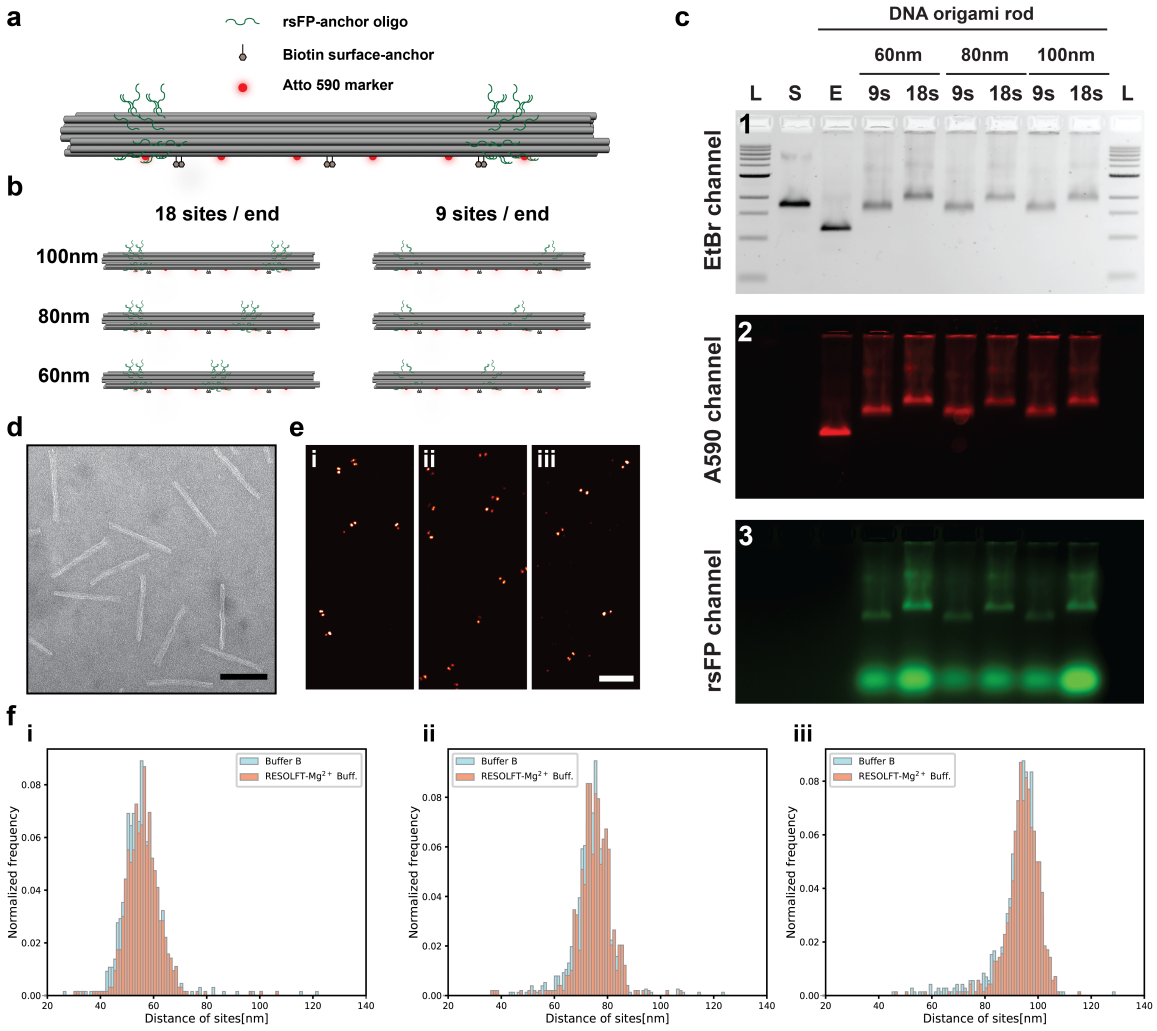

Figure 23. DNA origami nanorods used for RESOLFT resolution probing. (a) Schematic representation of the 140nm long DNA origami nanorod used in the experiments with cylinders representing double helices, highlighting the different functional groups on the structure. **(b)** Schematic representation of the different versions of the DNA origami nanorods used in experiments with anchoring regions containing 18 or 9 anchoring sites positioned at approximately 100 nm, 80 nm or 60 nm from each other. **(c)** rsFP functionalized DNA origami nanorods with 9 (9S) and 18 sites (18S) at 60 nm, 80 nm and 100 nm along with p7560 ssDNA scaffold (S), empty DNA origami nanorod (E) and DNA ladders (L) ran in 2% agarose gel stained with ethidium bromide and imaged with UV (1), red light (2) and green light (2) illumination. **(d)** Negative-stained TEM image of the purified empty DNA origami nanorods (scale bar = 100 nm). **(e)** Field of view DNA-PAINT super resolution images of DNA origami nanorods with DNA-PAINT docking strand functionalized rsFP-anchoring sites at **(e/i)** 60 nm, **(e/ii)** 80 nm and **(e/iii)** 100 nm in RESOLFT-Mg^{2+} Buffer (scale bar= 500 nm). **(f)** Plots of the calculated site distance distributions of structures with sites at **(f/i)** 60 nm, **(f/ii)** 80 nm and **(f/iii)** 100 nm from each other in Buffer B (cyan) and RESOLFT-Mg^{2+} Buffer (salmon).

4.3.2 Estimating functionalization yield of rsFP-modified DNA origami nanostructures

We than characterized the functionalization yield of rsFP modified nanostructures using a gel-based assay and quantitative DNA-PAINT (qPAINT) imaging described earlier. The gel-based assay showed a functionalization yield of around 60% for the nanorods and between 40% and 60% for the nanosheets. We prepared rsFP conjugates carrying DNA-PAINT docking sites in

order to carry out qPAINT experiments to look at the relative functionalization of sites on the nanorods within and between structures for the two proteins. From this data we found that the DNA nanorods functionalized with the two different proteins do not show different overall functionalization as the distribution of localization numbers, which scales with the number of proteins per structure, show almost perfect overlap. We also found that the relative functionalization of positions in the nanorods that can have an impact on the perceived resolution with RESOLFT microscopy was also nearly identical in the case of both rsFPs and showed that more than 60% of structures had higher than 0.6 relative functionalization. As the numbers of rsFPs per origami for the two types of structures were measured to be comparable and the samples did not show any systematic bias in terms of functionalization, we then used these well characterized samples for investigating the rsFPs photophysical behavior.

4.3.3 Quantitative characterization of rsFPs and the measurement of achievable resolution with DNA origami

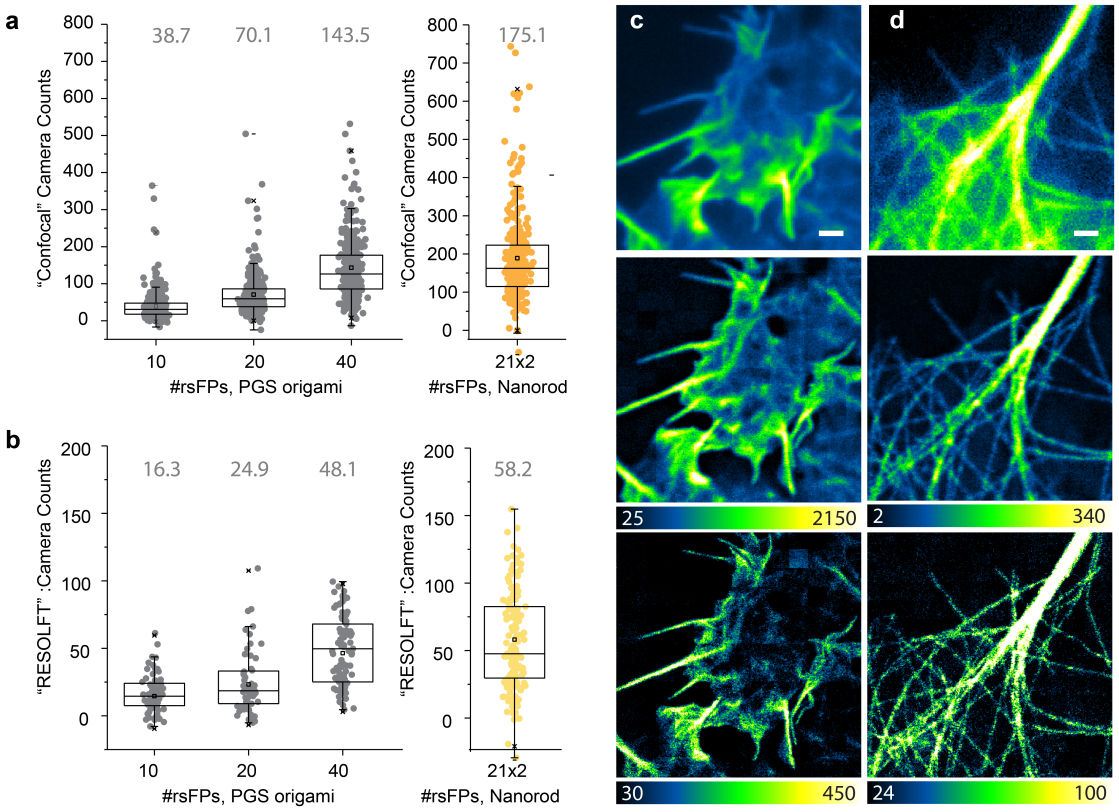

Figure 24. Calibration of the rsEGFP(N205S) labeled DNA origami structure photophysical behavior respect to the number of incorporated rsFPs. (a) Increase in the number of counts as a function of the increased number of available sites for rsEGFP(N205S) on the PGS structure. The counts are recorded in confocal mode for each identified spot and the mean over an area of 200x200 nm^2 is reported. Plots on the left show the counts measured for rsEGFP(N205S) functionalized DNA origami sheets with 10, 20, 40 sites, while the plot on the right show the count values for the rsEGFP(N205S) functionalized rod structure with 18-18 sites at a distance of 100 nm. **(b)** Same comparison in RESOLFT imaging mode. **(c-d)** Imaging of two different sub cellular structures, Lifeact-rsEGFP(N205S) and map2-rsEGFP2(N205S) respectively. The first row shows the wide-field image, the second the enhanced confocal and the last the REOLFT image. In the bottom count-intervals for the enhanced confocal and RESOLFT are reported. (scale bar = 1µm)

We then first immobilized rsEGFP(N205S) functionalized PGS 10S, 20S and 40S structures in homemade flow chambers to investigate the photophysical parameters (OFF- and ON-

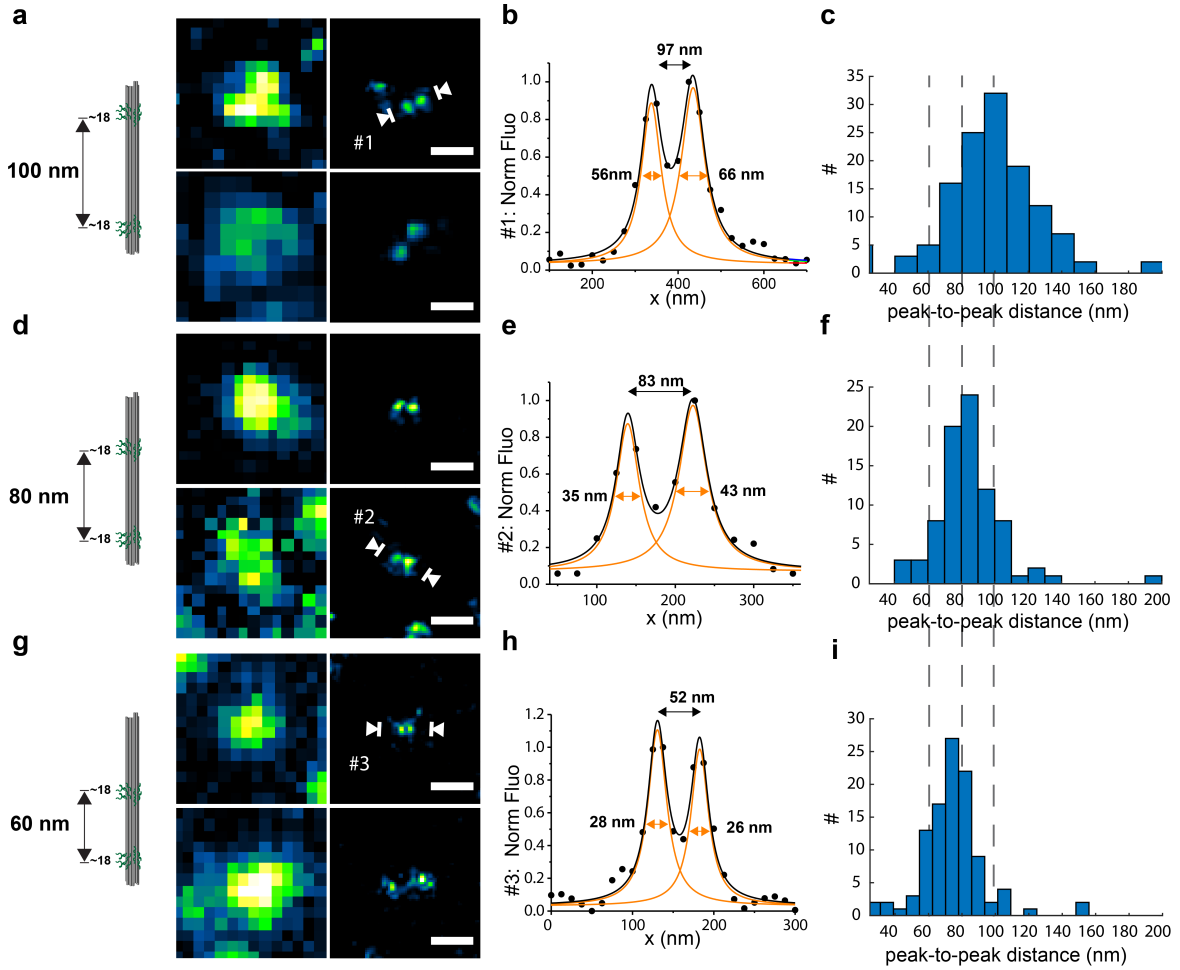

Figure 25. RESOLFT imaging of rsEGFP(N205S) labeled DNA origami nanorods. Wide-field (left) and super-resolution image (right) of single DNA origami structures with 18-18 addressable labeling-sites at approximately **(a)** 100 nm, **(d)** 80 nm and **(g)** 60 nm distance (scale bars= 250 nm). **(b, e, h)** Normalized intensity profiles measured across the lines (averaged over 50 nm) marked with white arrowheads in the super-resolved images. Each line profile has been fitted with the sum of two Lorentz function, the FWHM (in nm), as well as the peak-to-peak distance, is reported in the plot. **(c, f, i)**. Histogram of N = 172 **(c)**, 103 **(f)** and 143 **(i)** single DNA origami structures across a 40x40 μm^2 for the three distances.

switching kinetics) of the controlled quantities of the protein. We first investigated the OFF-switching halftime dependence of rsEGFP(N205S) on the 488 nm light intensity on the different PGS origami structures, which showed a trend that overlaps with the trend observed for tags in cell experiments. Another interesting observation we made was for the fluorescent counts we measured for PGS 40S structures at increasing ON-switching illumination (405 nm): the count values per DNA origami reached a saturation level, while the fatigue resistance of the proteins over the same 405 nm illumination range decreased, reaching a plateau value around 500 counts and this trend remained consistent for PGS 20S and PGS 10S structures as well. We also investigated the fluorescent count values and fatigue for the different amounts of rsFPs per structures and we observed that both these parameters scaled linearly with the number of rsFPs (Fig. 24) making these parameters useful for gaining quantitative information from RESOLFT data. We then used rsEGFP(N205S) functionalized nanorods to test the achievable resolution of RESOLFT. First, we used the nanorods with 18-18 sites at 60 nm, 80 nm and 100 nm from each other, which seemed to carry comparable number of rsFPs, as designed, to the PGS structures based on the measured fluorescent count values for the 100 nm structures (PGS 40S: 143.5 counts, 100 nm nanorod: 175.1 counts) (Fig. 25). We were also able to follow the

decrease of the distances over the defined range (Fig. 25) and we saw a similar trend in the case of rsEGFP2 functionalized nanorods as well. Finally, in the case of rsEGFP(N205S) we were also able to resolve rods carrying only 9 sites per end for the 100nm structures.

4.4 PAPER IV.

4.4.1 Production and characterization of DNA origami barcode library version 1

The high sensitivity of sequencing techniques made it possible to study gene expression on a single cell level. Another current aim of the transcriptomic field is to be able to link transcriptomics data to spatial position. There have been a number of impressive techniques developed, however these are still limited in either their spatial resolution or sequencing depth. The utilization of complex optical tagging systems together with microscopy for spatial transcriptomics have been attempted, but the DNA origami techniques' potential for creating such a labeling system has remained untapped.

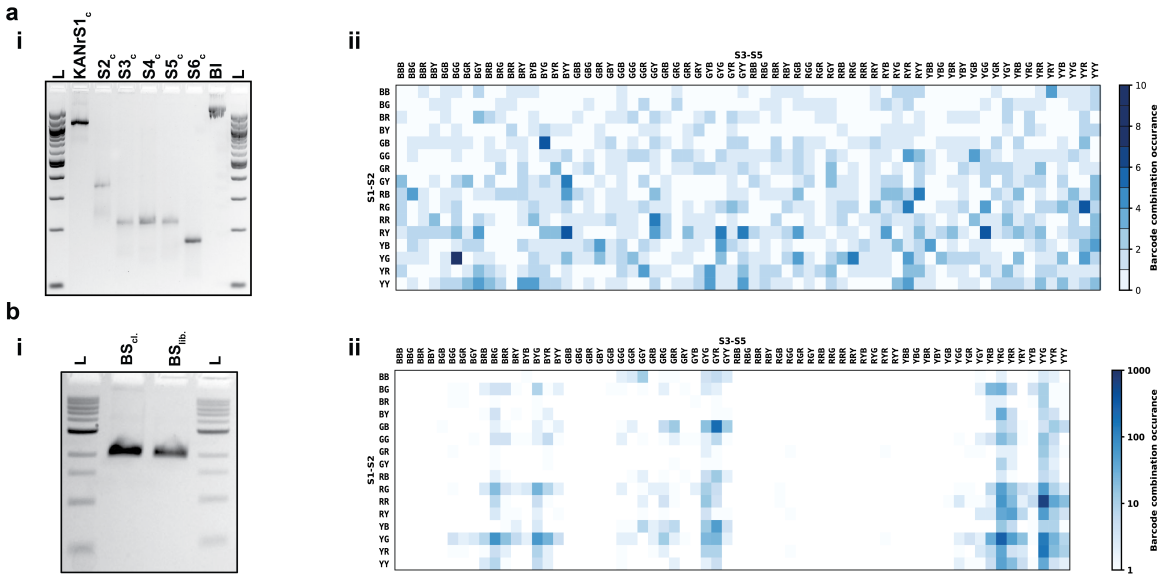

Figure 26. Production and NGS characterization of the barcode insert and scaffold library (v1). (a/i) Color spacers (KAN_RS1c - S6c) run in 0.8% agarose gel along the assembled barcode insert (BI) and DNA ladders (L). **(a/ii)** Heatmap representing the PacBio RS II NGS sequencing data of the insert library with each row representing the first two sites (S1-S2) and each column representing the last three sites (S3-S5) of a particular color combination. We identified 614 unique combinations and the frequency of occurance was observed to be within one order of magnitude (maximum frequency: 7, minimum frequency: 1). **(b/i)** DNA origami barcode scaffold library (v1) ($BS_{lib.}$) run along with DNA origami barcode scaffold from a clone ($BC_{cl.}$) and DNA ladders (L) in a 2% agarose gel. **(b/ii)** Heatmap representing the PacBio RS II NGS sequencing data of the scaffold library (v1). We identified 305 unique combinations and the frequency of occurrence showed a bias for certain combinations (maximum frequency: 681, minimum frequency: 1).

In paper IV. we developed a DNA origami-based tagging system for this purpose. For the first version of the library, as well as for the second version, we constructed a barcode insert library encoding the optical code of each barcode using a combinatorial PCR reaction using color spacer sequences, spacers modified with a mixture of color sites and overlapping sequences to the adjacent color sites, as described earlier (Fig. 26). After the assembly we characterized the insert library (v1) using PacBio RS II sequencing, providing long enough read length to cover

the full insert, and we observed near full coverage (614 unique combinations) despite the considerably low number of reads (5101 reads after QC filtering) with frequency of combinations lying within one order of magnitude compared to each other (maximum frequency: 7, minimum frequency: 1) (Fig. 26). The resulting insert library was then cloned into M13mp18, and the ssDNA scaffold library was produced as described in the methods section of this thesis. We characterized the scaffold library as well using PacBio RS II sequencing, and we observed a decreased coverage (305 unique combinations) in the data with the frequency distribution of the library showing a biased enrichment of certain combinations (maximum frequency: 681, minimum frequency: 1) despite the higher number of reads (9065 reads after QC filtering) (Fig. 26). In addition to this we discovered upon observing the frequency of color occurrence that a large number of color sites were grouped as ambiguous sites (sequences with high number of mismatches) or non-anticipated sites (sites with close similarity to non-encoded docking site combinations), whose frequency seemed to account to large extent for the higher observed frequency of the mixed "Y" site even in the case of the insert library.

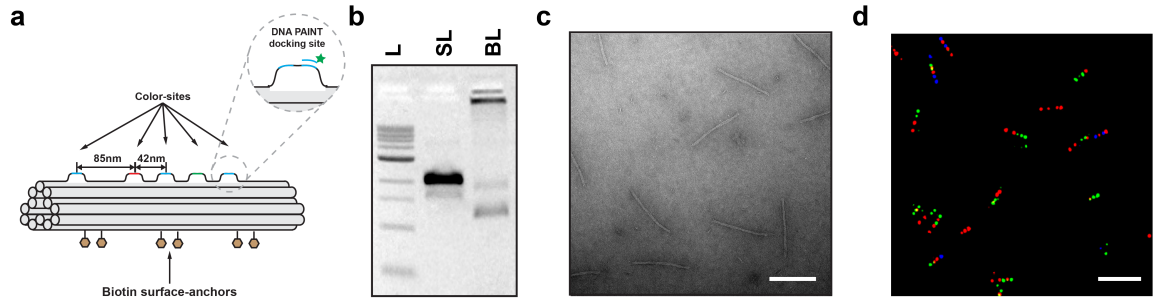

Figure 27. Folding and SR-imaging of DNA origami barcode library (v1) (a) Schematic representation of DNA origami barcode structure (v1) with five single-stranded color-site loops (carrying two DNA-PAINT docking sites) at 42 nm from each other with the first and second site being 85nm from each other to introduce asymmetry. The structure was designed to carry 6 biotinylated surface immobilization sites. **(b)** The scaffold library (SL) and the barcode library (BL) run in a 2% agarose gel along a DNA ladder (L). **(c)** TEM images of folded barcodes taken with 22000X magnification and (scale bar = 200 nm) **(d)** Field of view image of DNA origami barcodes (v1) immobilized in glass bottom flow chamber slides imaged with three-color Exchange-PAINT (scale bar = 500 nm).

We then tested the folding of the scaffold library into a barcode structure (Fig. 27) using a single set of staples. The folding was observed to be as successful as is expected for standard origami folding reactions (Fig. 27). We finally tested the barcode library with 3 color Exchange-PAINT SR imaging (Fig. 27). We observed 86 unique combinations out of 1473 detected barcodes in three 39 µm by 39 µm field of view image.

4.4.2 Production and characterization of DNA origami barcode library version 2

For the second version (v2) of the barcode library we implemented some changes to address shortcomings of the first library: we shortened the spacer to decrease the minimum read length required for the barcode library characterization to be able to use the higher throughput Illumina Miseq platform for it, we turned to using four distinct DNA-PAINT docking sequences in the color sites in order to avoid the problem of non-encoded sites observed in the case of the first

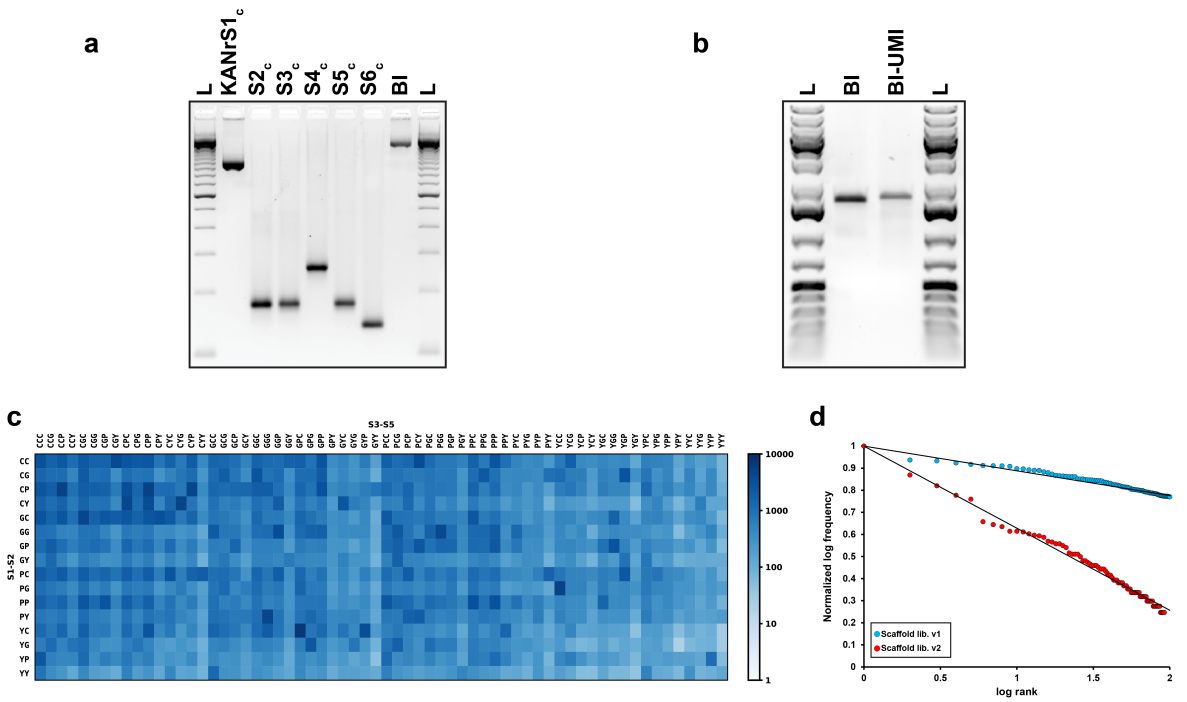

Figure 28. Production and NGS characterization of the barcode insert and scaffold library (v2). (a) Color spacers (KAN_RS1c-S6c) run in 0.8% agarose gel along the assembled barcode insert (BI) and DNA ladders (L). **(b)** UMI-labelled barcode insert library (BI-UMI) run in a 0.8% agarose gel along with unlabeled barcode insert library (BI) and DNA ladders (L). **(c)** Heatmap representation of the NGS data generated from sequencing the scaffold library v2 using the Illumina MiSeq platform with each row representing the first two sites (S1-S2) and each column representing the last three sites (S3-S5) of a particular color combination. We observed all of the designed combinations with the frequency of occurrence being within (maximum frequency: 7415, minimum frequency: 21). **(d)** Log-log plot of the ranked occurrence frequency of the color combinations in the scaffold library (v2) (blue) showing high homogeneity of the library compared to the similar plot for the scaffold library (v1) (red).

version of the library, we also devised a strategy utilizing unique molecular identifiers (UMIs) in order to compress the information of a barcode sequence into a tag readable by short read length platforms more commonly used for RNA sequencing and finally altered the cloning strategy to increase the coverage of the library and to decrease the bias in the barcode frequency distribution. We used a similar approach to assemble the KAN_R-fused insert library (Fig. 28) and added a ligation step before cloning to link a 25 nt UMI sequence to the assembled insert (Fig. 28). The resulting insert library was then cloned into M13mp18 using CPEC cloning, as described previously, and phage non-competent (F^-) bacterial strain was used to produce the scaffold library, in order to prevent the skewing of library distribution by exponential amplification of random differences between constructs during cloning. The ssDNA scaffold library was characterized after purification using paired-end sequencing on the Illumina MiSeq platform. We observed a full coverage of the designed library in the data with the frequency distribution of the color combinations showing large homogeneity compared to barcode library (v1) (Fig. 28). As for the UMI sequences, even though dominant UMI-sequences were observed to be linked to each color combination, UMI sequences still showed some promiscuity with low individual frequencies that was attributed to PCR hybridization.

We then tested the folding of the scaffold library (v2) into a barcode structure (Fig. 29) using a single set of staples. The folding was observed to be as successful as it was for barcode (v1) (Fig. 29). We then tested the barcode library with 4 color Exchange-PAINT SR imaging (Fig.

29). We observed 212 unique combinations out of 279 detected barcodes in a 44 µm by 44 µm field of view image.

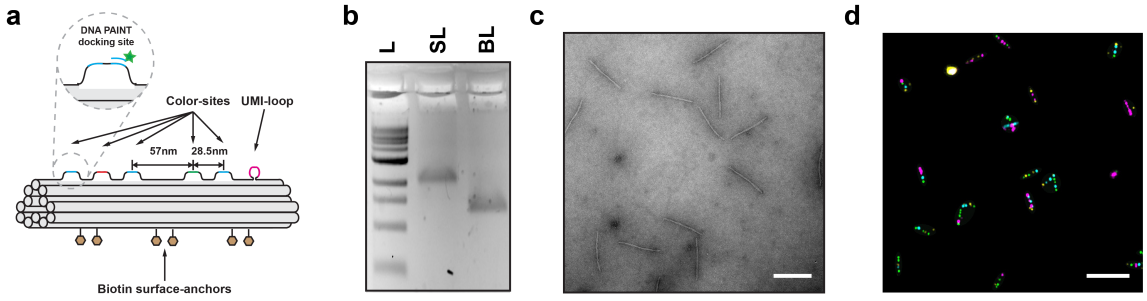

Figure 29. Folding and SR-imaging of DNA origami barcode library (v2) (a) Schematic representation of DNA origami barcode structure (v2) with a single-stranded loop containing the UMI sequence and five single-stranded color site loops (carrying two DNA-PAINT docking sites) at 28.5 nm from each other with the second and third site being 57 nm from each other to introduce asymmetry. The structure was designed to carry 6 biotinylated surface immobilization sites. **(b)** The scaffold library (SL) and the barcode library (BL) run in a 2% agarose gel along a DNA ladder (L). **(c)** TEM images of folded barcodes taken with 18000X magnification (scale bar = 200 nm) **(d)** Field of view image of DNA origami barcodes (v2) immobilized in glass bottom flow chamber slides imaged with four color Exchange-PAINT (scale bar = 500 nm).

4.4.3 Testing utilization of DNA origami barcodes

To test the applicability of these barcodes as an optical cell-tagging system for sequencing we prepared barcode structures with azide-anchors for cell functionalization. We then used NHS-PEG_4-DBCO crosslinker to functionalize cell surface proteins on fixed HeLa cells and were able to use the introduced DBCO groups to successfully attach azide-group carrying Cy5-labelled barcodes covalently to the cells surface by Copper-free click chemistry (Fig. 30). Finally for testing the compatibility of the barcodes with commonly used single cell RNAseq sample preparation protocols we prepared cDNA libraries from Fibroblast total RNA extracts containing decreasing amount of spiked in origami barcodes (800, 400 and 200) using a the SmartSeq2 protocol[98] (Fig. 30). We were able to detect the fragment amplified from the barcodes containing the UMI sequence at all of the dilutions along the cDNA library with the Bioanalyzer instrument (Fig. 31).

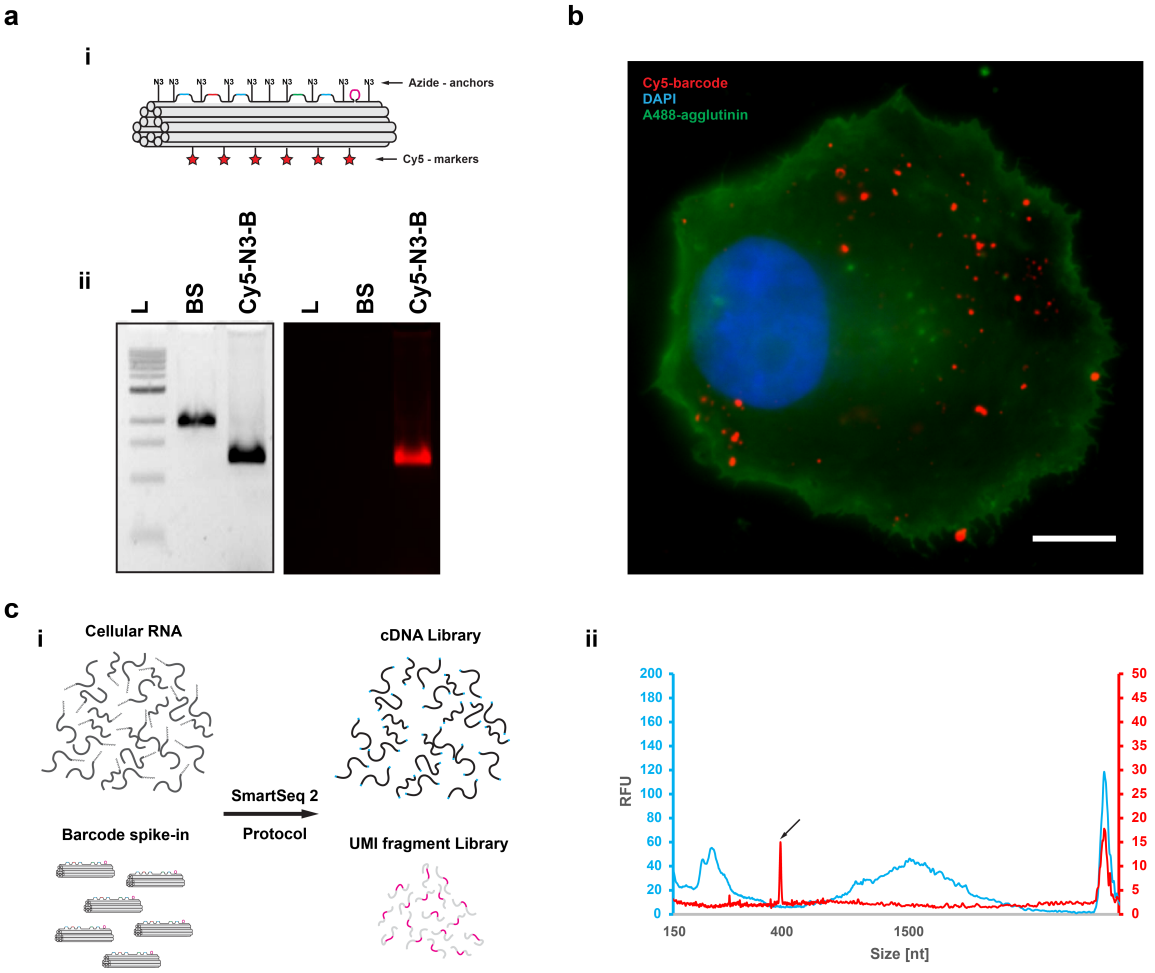

Figure 30. Utilization of DNA origami barcodes (v2) as cellular tags for scRNAseq (a/i) Schematic representation of DNA origami barcode structure used for functionalization of HeLa cells. The structure was designed to carry six Cy5 markers for visualization and ten azide-anchors used for attachment to cells. **(a/ii)** DNA origami barcode structures carrying Cy5 and azide groups (Cy5-N3-B) used for cell-functionalization experiment run in 2% agarose gel along with barcode scaffold (BS) and DNA ladders (L). **(b)** Fluorescent microscope image of a fixed HeLa cell with the nucleus stained with DAPI(blue), the plasma-membrane stained with Alexa Fluor 488-WGA (green) and functionalized with Cy5-labeled DNA origami barcodes (red) (scale bar = 10 µm). **(c/i)** Schematic representation of the experiment to coamplify mRNA cDNA with DNA origami barcode's (v2) UMI sequence from Fibroblast total RNA extract (50 pg) with 200 barcodes spiked in. **(c/ii)** Bioanalyzer traces of cDNA libraries (red) and DNA origami barcode UMI fragment (blue) amplified using the SmartSeq2 single cell RNA sequencing (scRNAseq) protocol from samples with 200 barcodes spiked in. The peak corresponding to the ~410 nt UMI fragment is highlighted with an arrow.

5 CONCLUSIONS

The papers presented in this thesis have further contributed to the existing body of work that highlights the DNA origami technique's ability to create predesigned molecular patterns with nanometer scale precision and exceptionally high yield and homogeneity. The variety of the scientific fields featured in this work also shows the universality of DNA origami as an experimental platform for creating research tools.

As it was demonstrated in paper I., biological systems are obvious candidates for targets as nanoscale patterns are known to play a central role in signaling pathways involved in a lot of biological phenomena such as immune reactions[99] or different morphogenetic processes[100]. The results presented in paper I. clearly suggest the role of nanoscale ligand distribution in activation of the probed ephrin-Eph signaling pathway. For other signaling pathways implicated in the before-mentioned biological processes the nanocaliper system could offer a relatively easily implementable approach to probe the effect of ligand distribution in a ligand concentration and surface decoupled way. A couple of potential avenues of improvement of this system could be the use of more specific ligand conjugation chemistries or the utilization of more complex 2D ligand patterns. Alternately the use of power generating structures[101] for probing pathways requiring mechanical input for activation (like the Notch signaling pathway[102]) or the use of structural features of the constructs to investigate alternative activation schemes (e.g. kinetic segregation model of TCR[103]) could be interesting possible research avenues for DNA origami based probes for cell signaling studies.

The continuous emergence and spreading of sub-diffraction resolution imaging techniques generated an ever-increasing need for reference samples with sub-diffraction sized features. As the work presented in paper II. and paper III. shows DNA origami provides an easily applicable way to assess a wide range of imaging methods utilizing fluorophores or fluorescent proteins. As we have shown, the precise programmability of these structures allows us to probe imaging methods for their experimentally achievable resolution or even faithfulness to the underlying reality of the sample. As we have presented in paper III., structures can also be made to be used as calibration samples to gain quantitative information from imaging data. The popular use of these structures as reference samples for imaging applications is also demonstrated by the fact that DNA origami-based reference structures for microscopy are now commercially available (from Gattaquant).

Apart from creating reference samples, another kind of imaging application for DNA origami structures, is the fabrication of multiplex optical molecular probes. In paper IV. we expanded on this idea with utilizing the so far unexplored approach of using the scaffold DNA's sequence to encode nanoscale patterns. With a combinatorial enzymatic approach, we created a DNA origami barcode library with high homogeneity and complexity that was demonstrated to be detectable by high-throughput NGS platforms as well as SR microscopy. We additionally showed that a UMI tagging strategy can be used to make the barcodes compatible with shorter read length sequencing platforms more commonly used in RNAseq workflows. Finally, we demonstrated the overall capability of the system to be used as a tagging system connecting the

methodologies of SR imaging and NGS by establishing the feasibility of functionalizing cells with origami barcodes and the inclusion of origami barcodes in the cDNA library preparation using the SmartSeq2 protocol.

6 ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, **Björn Högberg**, for giving me the opportunity to do research in this exciting field, for letting me explore things and for creating a lab with such an inspiring atmosphere.

I would also like to thank my colleagues who helped me through the tough times: Thank you **Mino** for all the tarallini and the daily Italian colloquial language lessons ☺. Thank you **Ioanna** for all the laughs and the constant supply of gummy snacks. Thank you**, Erik,** for answering all my stupid Swedish questions and for always helping me out even if it meant driving us home with our newborn babies from the hospital. Thank you, **Giulio**, for being a great friend, with whom I can share my scientific and private thoughts as well. Thank you, **Ian**, for helping me out with computing and math and for all the exciting, often philosophical discussions. I also want to thank the rest of my present and past colleagues **Andreas**, **Alan**, **Joao**, **Esther**, **Iris**, **Yang**, **Yunshi, Marco** and everyone else who I might have forgotten to mention for making it fun to come to work.

I would like thank my co-supervisor **Ana Teixeira** and the present and past members of her lab (**Vanessa**, **Alessandro**, **Elena**, **Christina**, **Trixy**, **Toon** and others) as well as the members of the **Stevens group** (**Chris**, **Hanna**, **Miina** and others) for all the helpful discussions and for making Biomaterials a fun Division.

I would like to thank my collaborators: **Hjalmar Brismar** and **Matthias Reuss** for introducing me to super resolution imaging, **Ilaria Testa** and **Francesca Pennacchietti** for teaching me about rsFPs and for all the help and enthusiasm they gave me, **Björn Reinius** and **Christos Coucoravas** for helping me out with everything related to RNA sequencing**.** I would also like to thank **Fritz Simmel** and **Ralf Jungmann** and the members of their labs for extensively helping me with DNA-PAINT imaging. And I want to also thank all the **members of the EscoDNA ITN** for all the nice meetings we had during those 3 years.

I want to thank all of my friends as well: Thank you **Theresa** and **Giulia** for all the nice times we had together!

I want to thank **my parents, Ferenc** and **Anikó**, for supporting me and for teaching me not to take anything on face value. I would also like to thank my **parents in law, Péter** and **Éva** for helping me and Kati in every possible way throughout the years we have known each other. Finally, I would like to thank my three adorable sons: **Ádám, Áron** and **Árvid** for always bringing a smile on my face even after a long, tiresome day at the lab. Lastly, I want to thank **my wonderful wife Kati**: I cannot describe how lucky I am to have you, you made all this possible.

7 REFERENCES

1. Miescher, F. Ueber die chemische Zusammensetzung der Eiterzellen (in German). *Med. Chem. Unters.* **4**, 441–460 (1871).

2. Kossel, A. Zur Chemie des Zellkerns (In German). *Zeitschrift für Physiol. Chemie* **7**, 7–22 (1883).

3. Avery, O. T. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *J. Exp. Med.* **79**, 137–158 (1944).

4. Neidle, S. *Principles of Nucleic Acid Structure*. (Elsevier, 2008). doi:10.1016/B978-0-12-369507-9.X5001-8

5. Burkholder, G. D., Latimer, L. J. & Lee, J. S. Immunofluorescent staining of mammalian nuclei and chromosomes with a monoclonal antibody to triplex DNA. *Chromosoma* **97**, 185–92 (1988).

6. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **12**, 847–854 (2005).

7. DiMaio, F. *et al.* A virus that infects a hyperthermophile encapsidates A-form DNA. *Science (80-.).* **348**, 914–917 (2015).

8. Wittig, B., Dorbic, T. & Rich, A. Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. *Proc. Natl. Acad. Sci.* **88**, 2259–2263 (1991).

9. Wang, G., Christensen, L. A. & Vasquez, K. M. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl. Acad. Sci.* **103**, 2677–2682 (2006).

10. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99**, 237–247 (1982).

11. Saiki, R. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science (80-.).* **239**, 487–491 (1988).

12. Consortium, I. H. G. S. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

13. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science (80-.).* **291**, 1304–1351 (2001).

14. Gibson, D. G. *et al.* Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science (80-.).* **319**, 1215–1220 (2008).

15. Carlson, R. The changing economics of DNA synthesis. *Nat. Biotechnol.* **27**, 1091–1094 (2009).

16. Matteucci, M. D. & Caruthers, M. H. Synthesis of deoxyoligonucleotides on a polymer support. *J. Am. Chem. Soc.* **103**, 3185–3191 (1981).

17. Kosuri, S. *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28**, 1295–1299 (2010).

18. Church, G. M., Gao, Y. & Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science (80-.).* **337**, 1628–1628 (2012).

19. Ducani, C., Kaul, C., Moche, M., Shih, W. M. & Högberg, B. Enzymatic production of 'monoclonal stoichiometric' single-stranded DNA oligonucleotides. *Nat. Methods* **10**, 647–652 (2013).

20. Praetorius, F. *et al.* Biotechnological mass production of DNA origami. *Nature* **552**, 84–87 (2017).

21. Kallenbach, N. R., Ma, R.-I. & Seeman, N. C. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **305**, 829–831 (1983).

22. Duckett, D. R., Murchie, A. I. & Lilley, D. M. The role of metal ions in the conformation of the four-way DNA junction. *EMBO J.* **9**, 583–90 (1990).

23. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).

24. Yan, H. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science (80-.).* **301**, 1882–1884 (2003).

25. Chen, J. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).

26. Goodman, R. P. Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science (80-.).* **310**, 1661–1665 (2005).

27. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science (80-.).* **338**, 1177–1183 (2012).

28. Ong, L. L. *et al.* Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components. *Nature* **552**, 72–77 (2017).

29. Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372 (2011).

30. Cherry, K. M. & Qian, L. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature* **559**, 370–376 (2018).

31. Bath, J. & Turberfield, A. J. DNA nanomachines. *Nat. Nanotechnol.* **2**, 275–284 (2007).

32. Gu, H., Chao, J., Xiao, S.-J. & Seeman, N. C. A proximity-based programmable DNA nanoscale assembly line. *Nature* **465**, 202–205 (2010).

33. Choi, H. M. T. *et al.* Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nat. Biotechnol.* **28**, 1208–1212 (2010).

34. Kick, B., Praetorius, F., Dietz, H. & Weuster-Botz, D. Efficient Production of Single-Stranded Phage DNA as Scaffolds for DNA Origami. *Nano Lett.* **15**, 4672–4676 (2015).

35. Nafisi, P. M., Aksel, T. & Douglas, S. M. Construction of a novel phagemid to produce custom DNA origami scaffolds. *Synth. Biol.* **3**, (2018).

36. Högberg, B., Liedl, T. & Shih, W. M. Folding DNA origami from a double-stranded

source of scaffold. *J. Am. Chem. Soc.* **131**, 9154–5 (2009).

37. Wang, P., Hyeon Ko, S., Tian, C., Hao, C. & Mao, C. RNA–DNA hybrid origami: folding of a long RNA single strand into complex nanostructures using short DNA helper strands. *Chem. Commun.* **49**, 5462 (2013).

38. Erkelenz, M. *et al.* A Facile Method for Preparation of Tailored Scaffolds for DNA-Origami. *Small* **10**, 73–77 (2014).

39. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).

40. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).

41. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into Twisted and Curved Nanoscale Shapes. *Science (80-.).* **325**, 725–730 (2009).

42. Andersen, E. & Nielsen, M. DNA origami design of 3D nanostructures. *Protoc. Exch.* (2009). doi:10.1038/nprot.2009.75

43. Marras, A. E., Zhou, L., Su, H.-J. & Castro, C. E. Programmable motion of DNA origami mechanisms. *Proc. Natl. Acad. Sci.* **112**, 713–718 (2015).

44. Gerling, T., Wagenbauer, K. F., Neuner, A. M. & Dietz, H. Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components. *Science (80-.).* **347**, 1446–1452 (2015).

45. Tikhomirov, G., Petersen, P. & Qian, L. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature* **552**, 67–71 (2017).

46. Wagenbauer, K. F., Sigl, C. & Dietz, H. Gigadalton-scale shape-programmable DNA assemblies. *Nature* **552**, 78–83 (2017).

47. Ke, Y. *et al.* Multilayer DNA Origami Packed on a Square Lattice. *J. Am. Chem. Soc.* **131**, 15903–15908 (2009).

48. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444 (2015).

49. Zhang, F. *et al.* Complex wireframe DNA origami nanostructures with multi-arm junction vertices. *Nat. Nanotechnol.* **10**, 779–784 (2015).

50. Veneziano, R. *et al.* Designer nanoscale DNA assemblies programmed from the top down. *Science (80-.).* **352**, 1534–1534 (2016).

51. Benson, E. *et al.* Computer-Aided Production of Scaffolded DNA Nanostructures from Flat Sheet Meshes. *Angew. Chemie Int. Ed.* **55**, 8869–8872 (2016).

52. Benson, E. *et al.* Effects of Design Choices on the Stiffness of Wireframe DNA Origami Structures. *ACS Nano* **12**, 9291–9299 (2018).

53. Talley, K. & Alexov, E. On the pH-optimum of activity and stability of proteins. *Proteins Struct. Funct. Bioinforma.* n/a-n/a (2010). doi:10.1002/prot.22786

54. Grossi, G., Dalgaard Ebbesen Jepsen, M., Kjems, J. & Andersen, E. S. Control of enzyme reactions by a reconfigurable DNA nanovault. *Nat. Commun.* **8**, 992 (2017).

55. Li, S. *et al.* A DNA nanorobot functions as a cancer therapeutic in response to a molecular trigger in vivo. *Nat. Biotechnol.* **36**, 258–264 (2018).

56. Fu, J., Liu, M., Liu, Y., Woodbury, N. W. & Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **134**, 5516–5519 (2012).

57. Schlichthaerle, T. *et al.* Site-Specific Labeling of Affimers for DNA-PAINT Microscopy. *Angew. Chemie Int. Ed.* **57**, 11060–11063 (2018).

58. Rosen, C. B. *et al.* Template-directed covalent conjugation of DNA to native antibodies, transferrin and other metal-binding proteins. *Nat. Chem.* **6**, 804–809 (2014).

59. Bernardinelli, G. & Högberg, B. Entirely enzymatic nanofabrication of DNA–protein conjugates. *Nucleic Acids Res.* **45**, e160–e160 (2017).

60. Derr, N. D. *et al.* Tug-of-War in Motor Protein Ensembles Revealed with a Programmable DNA Origami Scaffold. *Science (80-.).* **338**, 662–665 (2012).

61. Hallam, T. J., Wold, E., Wahl, A. & Smider, V. V. Antibody Conjugates with Unnatural Amino Acids. *Mol. Pharm.* **12**, 1848–1862 (2015).

62. Janeway Jr, C. A., Paul, T., Mark, W. & Shlomchik, M. J. *Immunobiology: The Immune System in Health and Disease.* (Garland Science, 2001).

63. Shaw, A. *et al.* Binding to nanopatterned antigens is dominated by the spatial tolerance of antibodies. *Nat. Nanotechnol.* **14**, 184–190 (2019).

64. Zhao, Y.-X. *et al.* DNA origami delivery system for cancer therapy with tunable release properties. *ACS Nano* **6**, 8684–91 (2012).

65. Jiang, Q. *et al.* DNA origami as a carrier for circumvention of drug resistance. *J. Am. Chem. Soc.* **134**, 13396–403 (2012).

66. Douglas, S. M., Bachelet, I. & Church, G. M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science (80-.).* **335**, 831–834 (2012).

67. Gerling, T., Kube, M., Kick, B. & Dietz, H. Sequence-programmable covalent bonding of designed DNA assemblies. *Sci. Adv.* **4**, eaau1157 (2018).

68. Ponnuswamy, N. *et al.* Oligolysine-based coating protects DNA nanostructures from low-salt denaturation and nuclease degradation. *Nat. Commun.* **8**, 15654 (2017).

69. Perrault, S. D. & Shih, W. M. Virus-Inspired Membrane Encapsulation of DNA Nanostructures To Achieve In Vivo Stability. *ACS Nano* **8**, 5132–5140 (2014).

70. Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch. für mikroskopische Anat.* **9**, 413–418 (1873).

71. Moran, U., Phillips, R. & Milo, R. SnapShot: Key Numbers in Biology. *Cell* **141**, 1262–1262.e1 (2010).

72. Hell, S. W. & Wichmann, J. Breaking the Diffraction Resolution Limit By Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt. Lett.* **19**, 780–782 (1994).

73. Heintzmann, R. & Cremer, C. G. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. in *Proc. SPIE 3568, Optical Biopsies and Microscopic Techniques III* (eds. Bigio, I. J., Schneckenburger, H., Slavik, J., Svanberg, K. & Viallet, P. M.) 185–196 (1999). doi:10.1117/12.336833

74. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**, 82–87 (2000).

75. Hofmann, M., Eggeling, C., Jakobs, S. & Hell, S. W. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl. Acad. Sci.* **102**, 17565–17569 (2005).

76. Rust, M. J., Bates, M. & Zhuang, X. W. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**, 793–795 (2006).

77. Heilemann, M. *et al.* Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angew. Chemie - Int. Ed.* **47**, 6172–6176 (2008).

78. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–5 (2006).

79. Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. *Biophys. J.* **91**, 4258–4272 (2006).

80. Jungmann, R. *et al.* Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Lett.* **10**, 4756–61 (2010).

81. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313–8 (2014).

82. Wade, O. K. *et al.* 124-Color Super-resolution Imaging by Engineering DNA-PAINT Blinking Kinetics. *Nano Lett.* **19**, 2641–2646 (2019).

83. Jungmann, R. *et al.* Quantitative super-resolution imaging with qPAINT. *Nat. Methods* **13**, 439–442 (2016).

84. Schmied, J. J. *et al.* Fluorescence and super-resolution standards based on DNA origami. *Nat. Methods* **9**, 1133–1134 (2012).

85. Schmied, J. J. *et al.* DNA origami nanopillars as standards for three-dimensional superresolution microscopy. *Nano Lett.* **13**, 781–785 (2013).

86. Jusuk, I., Vietz, C., Raab, M., Dammeyer, T. & Tinnefeld, P. Super-Resolution Imaging Conditions for enhanced Yellow Fluorescent Protein (eYFP) Demonstrated on DNA Origami Nanorulers. *Sci. Rep.* **5**, 14075 (2015).

87. Lin, C. *et al.* Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA. *Nat. Chem.* **4**, 832–839 (2012).

88. Woehrstein, J. B. *et al.* Sub–100-nm metafluorophores with digitally tunable optical properties self-assembled from DNA. *Sci. Adv.* **3**, e1602128 (2017).

89. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001–5006 (2009).

90. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).

91. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).

92. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6**, 242–251 (2011).

93. Cong, Y. *et al.* Site-Specific PEGylation at Histidine Tags. *Bioconjug. Chem.* **23**, 248–263 (2012).

94. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

95. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

96. Macrae, M. *et al.* A conditional feedback loop regulates Ras activity through EphA2. *Cancer Cell* **8**, 111–118 (2005).

97. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).

98. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

99. Manz, B. N., Jackson, B. L., Petit, R. S., Dustin, M. L. & Groves, J. T-cell triggering thresholds are modulated by the number of antigen within individual T-cell receptor clusters. *Proc. Natl. Acad. Sci.* **108**, 9089–9094 (2011).

100. Gilboa, L. *et al.* Bone Morphogenetic Protein Receptor Complexes on the Surface of Live Cells: A New Oligomerization Mode for Serine/Threonine Kinase Receptors. *Mol. Biol. Cell* **11**, 1023–1035 (2000).

101. Liedl, T., Högberg, B., Tytell, J., Ingber, D. E. & Shih, W. M. Self-assembly of three-dimensional prestressed tensegrity structures from DNA. *Nat. Nanotechnol.* **5**, 520–524 (2010).

102. Chowdhury, F. *et al.* Defining Single Molecular Forces Required for Notch Activation Using Nano Yoyo. *Nano Lett.* **16**, 3892–3897 (2016).

103. Davis, S. J. & van der Merwe, P. A. The kinetic-segregation model: TCR triggering and beyond. *Nat. Immunol.* **7**, 803–809 (2006).

104. Williams, R. *et al.* Amplification of complex gene libraries by emulsion PCR. *Nat. Methods* **3**, 545–550 (2006).

105. Shagin, D. A. *et al.* Application of nonsense-mediated primer exclusion (NOPE) for preparation of unique molecular barcoded libraries. *BMC Genomics* **18**, 440 (2017).